

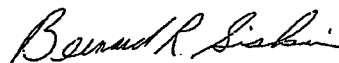
EXHIBIT V

REPORT

in the matter of

United States of America v. City of New York

by



Bernard R. Siskin, Ph.D.
Director and Head of Labor Practice Group

LECG
Philadelphia PA

July 2008

-5-

Exams 7029 or 2043 measure different, distinguishable abilities as the City intended, then items that measure the same ability (*e.g.*, the 12 items that measure Problem Sensitivity) should relate more closely to each other than to items measuring different abilities. To the extent that items may measure a common construct (*e.g.*, general intelligence), items also will be correlated across abilities to some degree but, since the items are purportedly designed to measure specific, discrete abilities, they should be much more highly correlated with other items measuring the same ability. That is, one would expect the items intended to measure the same ability to be more highly correlated with each other than with items intended to measure other abilities.²

8. The average correlations between items intended to measure the same ability and items intended to measure different abilities are presented in Tables 1 and 2 (for Written Exams 7029 and 2043, respectively). In this context, the correlation coefficient is a measure of the relationship between test items. Correlation values range from 0 (indicating no relationship) to 1 (indicating perfect relationship). Items that measure the same ability should be highly correlated. Items that measure different abilities should be correlated only to the extent that they are both affected by some common factor, such as general intelligence, and, thus, should have a lower correlation. Tables 1 and 2 show the correlations for each of the item sets representing the nine cognitive abilities. Looking at any column in these tables, the first value reported is the average correlation between items intended to measure the same cognitive ability. For example, on Table

² See Nunnally, J. C., (1978). *Psychometric Theory*, pp.103-05. New York: McGraw-Hill; Guion, R. M., (1998). *Assessment, Measurement, and Prediction for Personnel Decisions*, p. 185. Mahwah, NJ: Erlbaum. Dr. Catherine Cline, who (with one of the City's expert witnesses, Dr. Bobko) developed Written Exam 6019, agrees. See Dep. Test. of Catherine Cline, March 24, 2008, p. 322; Dep. Test. of Philip Bobko, July 1, 2008, pp. 188-89 (saying that to determine whether the written exams succeeded in measuring each of the nine abilities, Dr. Bobko would see whether items measuring the same ability correlated with one another).

-6-

1; we see that the set of items intended to measure Deductive Reasoning on Written Exam 7029 had an average correlation to each other (*i.e.*, a within-ability correlation) of 0.139. Table 1 further shows that those same Deductive Reasoning items correlated 0.136, on average, with the items intended to measure Inductive Reasoning, and correlated 0.135, on average, with the items intended to measure Information Ordering. Looking at all the values in the Deductive Reasoning column of Table 1, we see a higher average correlation between the items intended to measure Deductive Reasoning and those intended to measure Written Expression (0.153), Written Comprehension (0.148) and Spatial Orientation (0.147) than was found between the items intended to measure Deductive Reasoning itself (0.139). In other words, for Deductive Reasoning, there are higher average across-ability correlations than average within-ability correlations.

9. Inspection of Table 1 shows that this sort of pattern seems to be common. That is, items intended to measure an individual cognitive ability actually tend to correlate as or more highly with items intended to measure different cognitive abilities than they do with items intended to measure the same cognitive ability. Looking at Table 1, everything except the items intended to measure Spatial Orientation correlate most highly with Written Expression. For Table 2, the pattern is similar but not so stark. Four of the nine abilities have items that correlate on average more highly with items intended to measure different abilities than with items intended to measure the same ability. The average within-ability correlations are just barely higher than the average across-ability correlations for the other five abilities.³ For example, the

³ That is, the five abilities that actually have higher within-ability correlations than the across-ability correlations.

-7-

average correlation between Inductive Reasoning items is .222, but the average correlation between Inductive Reasoning items and Information Ordering items is .215. Likewise, the correlation for Memorization items to other Memorization items is .087, but the correlation for Memorization items to Visualization items is .080. In both cases there is a difference in correlation of only .007.

10. This pattern clearly demonstrates that the items on Written Exams 7029 and 2043 do not measure nine distinct abilities, as they were designed to do. Thus, statistical evidence shows that the written examinations fail to measure and weight the nine ability constructs consistent with what the test developers' job analysis deemed to be relevant to performance.

11. Another, more formalized approach to inspecting the patterns of relationships between items relative to the test plan is the statistical method known as factor analysis. Factor analysis is a statistical methodology that, based on the empirical data, defines an underlying structure which can explain the correlations among the items. Specifically, factor analysis (i) identifies unique, independent factors (defined by weighted combinations of the examination items) which explain the observed pattern of correlation among the items, and (ii) calculates the percent of the shared variance among the items (*i.e.*, common variance) that can be explained by each factor. In the context of examining the relationship between the items on Written Exams 7029 and 2043, the factors refer to the different groupings of items based on their interrelationships. Weights (or "loadings") are assigned to each item, indicating the correlation between that item and the factor.

12. For the results of factor analysis to confirm the test plan, the analysis should find

-8-

that items group together to comprise nine or 10 factors in a manner consistent with the test plan,⁴ such that the Deductive Reasoning items group together to form one factor and the items intended to measure Inductive Reasoning group together to form a second factor, and so forth. The factor analyses for Written Exams 7029 and 2043 show no such result.

13. The results of the factor analyses are reported on Tables 3 and 4 (for Written Exams 7029 and 2043, respectively).⁵ The analysis revealed a large primary factor and a smaller secondary factor for both Written Exams 7029 and 2043.⁶ For Written Exam 7029, the primary factor accounted for 80 percent of the common variance; a secondary factor added 10 percent for a total percentage of common variance accounted for by the two factors of 90 percent. For Written Exam 2043, the primary factor accounted for 73 percent of the common variance and a secondary factor added 15 percent for a total percentage of common variance accounted for by the two factor solution of 88 percent. Hence, the two empirically determined factors (*i.e.*, ability domains) explain approximately 90 percent of the pattern of the correlation among the items. The remaining 10 percent is "explained" by the unique measurement characteristics of the individual items, and not by some constructed cognitive ability domain.

14. Tables 3 and 4 show (i) the item number from the A.M. administration for both

⁴ In addition to nine factors corresponding to the nine discrete abilities, a single common factor may also be expected to measure a general cognitive ability (*i.e.*, general intelligence) that would influence all the nine discrete abilities.

⁵ The method used was the most commonly accepted method – principal components factor analysis with orthogonal varimax rotation.

⁶ The conventional requirement of minimum Eigen-value of 1 for factor retention was employed.

-9-

written examinations, (ii) the intended cognitive ability or factor for the item, and (iii) the factor loading of the items on the primary and the secondary factors.

15. While factors are empirically determined, the interpretation of what a factor represents is subjective.⁷ The factor analysis results summarized in Tables 3 and 4 show that the written examinations seem to primarily measure a general cognitive ability (except, perhaps, Memorization), and to a much lesser extent, a second specific cognitive ability (which is different from any defined by the test developers). The factor analysis clearly does not demonstrate that the goal of the test developers was met. The data does not factor into nine distinct factors or ability domains. This result demonstrates that the purported intent of the test design (to measure and weight nine distinct cognitive ability domains) was not successful.⁸

16. Considering both the correlations describing the relationship patterns between the items and the factor analyses, one finds no support for the claim that Written Exams 7029 and 2043 measure nine different cognitive ability domains. To the extent that the City's claim of validity of the written examinations is based on the linkages or importance ratings of nine cognitive ability domains to different job tasks as established in the job analysis work reported in the Morrongiello Report, the available statistical evidence indicates that Written Exams 7029 and

⁷ In order to interpret the domain that each factor represents, one commonly assigns items (in a mutually exclusive fashion) as best characterizing a specific factor if (i) the item has a factor loading of at least .30 on that factor and (ii) that loading is at least .10 higher than that item's factor loading on any other factors. See Nunnally, 1978. *Psychometric Theory*, p.423.

⁸ I also forced a nine factor solution by eliminating the minimum Eigen value constraint for defining a factor, to see whether the nine forced factors would correspond to the nine ability domains the test was designed to measure. They did not. With the exception of Memorization and, for Written Exam 2043, Spatial Orientation, none of the nine forced factors represented the ability domains the test developers purportedly designed the test to measure. Moreover, the importance of the nine factors in no way corresponds to the purported intent of the test design.

-10-

2043 do not measure the nine discrete abilities that were rated important for firefighter job performance.

ISSUE 2(a): THERE IS NO JUSTIFICATION FOR THE CUTOFF SCORES USED FOR WRITTEN EXAMS 7029 AND 2043.

17. The City used a cutoff score of 84.705 for Written Exam 7029 and a cutoff score of 70.00 for Written Exam 2043. As I demonstrated in my prior report, the pass/fail use of Written Exams 7029 and 2043 with these cutoff scores has resulted in a statistically significant disparate impact upon both African-American and Hispanic candidates. The City has stated that it choose the cutoff scores for administrative reasons,⁹ but it has offered no psychometric justification. The City does not contend that the cutoff scores used for Written Exams 7029 and 2043 measure the minimum level of the abilities purportedly tested that is necessary for successful performance of the job of entry-level firefighter in the FDNY.¹⁰

18. The City's experts, Drs. Philip Bobko and F. Mark Schemmer, appear to claim, based on meta-analysis, that (i) performance on cognitive tests correlates with firefighter job

⁹ For example, the City states that it chose 84.705 as the cutoff score on Written Exam 7029 so that there would be a sufficient number of candidates on the Exam 7029 eligibility list, based on the number of hires the FDNY expected to make over the life of the list, but not so many that a large number of candidates on the list would not be reached for hire before the list expired. *See* Dep. Test. of Carol Wachter, January 17, 2008, pp. 74-75, 86-88; Dep. Test. of Sherry Kavalier, February 22, 2008, p. 384-85; Rule 30(b)(6) Dep. Test. of the City (T. Patitucci, designee), pp. 90-91, 102-103; *see also* Personnel Rules and Regulations for the City of New York, Rule 4.4.9 (setting the default pass mark on an examination at 70).

¹⁰ *See* City's Resp. to Interrog. No. 30 in Def.'s Supplemental Resps. and Objections to Pl.'s United States' Interrogs. (Nos. 29-31). The City also does not contend that there is any score or range of scores on Written Exam 7029 or 2043 that corresponds to the minimum level of cognitive ability necessary to perform the job successfully. *See* the City's Resps. to Interrogs. Nos. 34 and 35 in Def.'s Resps. and Objections to Pl.'s First Set of Reqs. for Admis. (Nos. 1-23), Second Set of Interrogs. (Nos. 33-35) and Second Set of Reqs. for Prod. of Docs. (Nos. 98-11).

-15-

Candidate	Eligible to be Hired?	Test Scores		Transformed Score ¹⁸ Exam 7029
		Written Exam 7029	Physical Exam 7029	
A	Yes	84.705	75.0	70.000
B	No	70.588 ¹⁹	100.0	70.813
C	Yes	90.353	87.5	83.024
D	No	83.529	100.0	83.652

As the chart above shows, the candidate (Candidate D) who is predicted by the examination results to perform best as a firefighter is not eligible to be hired, because he failed Written Exam 7029 at the cutoff score of 84.705 used by the City. Conversely, the candidate (Candidate A) who is predicted by the examination results to perform worst as a firefighter is eligible to be hired. Exclusion of individuals based solely on their written score cannot be justified by the assumption that the written examination is correlated with performance (*i.e.*, "more is better") since (i) there is no evidence that those who scored below 84.705 on the written examination do not have the necessary skills, abilities and other characteristics to do the job and, more significantly, (ii) if those who failed the written examination had been allowed to take the PPT, they may have been predicted to be better performers than those actually hired, when their written and physical skills and abilities were jointly assessed.

ISSUE 2(b): THE SCORING METHODOLOGY FOR WRITTEN EXAM 6019 INDICATES THAT THE CUTOFF SCORES ON WRITTEN EXAMS 7029 AND 2043 WERE TOO HIGH.

¹⁸ The transformed scores were calculated using the methodology set forth in the Firefighter Exam. No. 7029 Explanation of Test Scores, Pl.'s Dep. Ex. 35. See Appendix D.

¹⁹ This is the minimum score on Written Exam 7029 which is at or exceeds 70 percent, the cutoff used on Written Exam 2043.

-16-

25. As noted above, the City has developed a new written examination, Written Exam 6019, which measures cognitive abilities and other important attributes not measured by Written Exams 7029 and 2043. Written Exam 6019 consists of three components: (i) timed, (ii) situational judgement, and (iii) cognitive abilities.²⁰ None of the components are scored as pass/fail. A candidate passes or fails Written Exam 6019 on the basis of whether or not the candidate's "final score" is at least 70 percent.²¹ See Firefighter, Exam No. 6019 – Explanation of Test Scores, Pl.'s Dep. Ex. 203, Appendix F; see also Dep. Test. of Catherine Cline, June 19, 2008, pp. 73-74. Therefore, poor performance on one component can be offset by good performance on the other components. On Written Exam 6019, the standardized cognitive score receives a weight of 40 out of 90 or 44.4 percent; the standardized timed score receives a weight of 15 out of 90 or 16.7 percent; and the standardized situational judgment score receives a weight of 35 out of 90 or 38.9 percent.²² The combined score is converted to a final average, and a final score of at least 70 is needed to pass Written Exam 6019.²³ See Firefighter, Exam No. 6019 –

²⁰ The timed component consists of items designed to measure memorization, speed of closure and perceptual speed. The situational judgment component consists of items designed to measure adaptability, tenacity, work standards, resiliency, coordination, integrity and interpersonal relationships. The cognitive component consists of items designed to measure deductive reasoning, flexibility of closure, information ordering, inductive reasoning, number facility, problem sensitivity, spatial orientation and visualization.

²¹ This is a transformation of the weighted average of the standardized score of each of the three components.

²² A candidate who scores at the mean of each test receives a combined weighted z score of 0, while one who scores one standard deviation above the mean on the cognitive component and at the mean on the other components would score 0.444 $((0*15+0*35+1*40)/90)$.

²³ The final average is computed as $9.817 * \text{the combined weighted z score} + 86.826$. In order to score at least 70 and pass Written Exam 6019, a candidate's combined weighted z score must be equal to or higher than -1.713.

-17-

Explanation of Test Scores, Pl.'s Dep. Ex. 203, Appendix F; *see also* Firefighter, Exam. No. 6019 Final Answer Keys, Pl.'s Dep. Ex. 202, Appendix G. A candidate who answered all of the cognitive component questions incorrectly (which would yield a negative cognitive component final score) could still pass the test if the candidate answered all of the other questions correctly; in that case, the candidate's final average would be 75.811, a passing score. This would indicate that the cognitive abilities can be totally offset by other skills and no cutoff would be appropriate for the cognitive component.

26. While there is no cutoff score on the cognitive component of Written Exam 6019, there was a cutoff score for Written Exams 7029 and 2043, which tested only cognitive abilities. Memorization was considered to be a cognitive ability in the construction of Written Exams 7029 and 2043, but it is part of the timed component of Written Exam 6019. To study the effect of this difference, I computed what would happen to the scores on Written Exam 6019 if a candidate answered all of the cognitive component items as well as all the Memorization items incorrectly, but answered all the other items on Written Exam 6019 correctly. My computations revealed that, even in that case, a candidate would still pass Written Exam 6019, with a score of 71.953. Similarly, another difference between Written Exams 7029 and 2043 and Written Exam 6019 is that Written Exams 7029 and 2043 were designed to measure directly the cognitive ability of Written Comprehension (9 of the 85 questions, or 10.6 percent of the questions on Written Exams 7029 and 2043 were designed to measure Written Comprehension), while Written Exam 6019 measured Written Comprehension only indirectly.²⁴ If we assume that poor

²⁴ According to Dr. Cline, Written Exam 6019 measures Written Comprehension indirectly in that one obviously needs some ability to comprehend written materials to understand the written questions asked in any written examination. *See* Dep. Test. of Catherine Cline,

-18-

Written Comprehension would reduce the number of non-cognitive questions that one would answer correctly by roughly 10 percent (actually 10.6 percent), then if one answers roughly 90 percent (actually 89.4 percent) of the non-cognitive questions correctly, one would only need to answer 31 percent of the cognitive questions correctly in order to pass. Thus, the scoring methodology for Written Exam 6019 indicates that if a cutoff score is appropriate on Written Exams 7029 and 2043, the cutoff scores on Written Exams 7029 and 2043 should have been well below 70 percent, and perhaps as low as 30 percent.

ISSUE 3: MOREOVER, THERE IS AFFIRMATIVE STATISTICAL EVIDENCE THAT THE CUTOFF SCORES THE CITY USED FOR WRITTEN EXAMS 7029 AND 2043 ARE NOT JUSTIFIED.

27. As discussed above, there is no valid statistical evidence that the cutoff scores the City used on Written Exams 7029 and 2043 are justified. However, I also was asked to determine whether there is any affirmative statistical evidence that the cutoff scores are not justified. That is, I was asked to determine to what extent there is evidence in the data provided by the City that those who failed either Written Exam 7029 or 2043 could nevertheless do the job of firefighter. Given the City's selection practices, it is reasonable to assume that passing either of the written examinations is an indicator that a candidate had the minimum cognitive ability necessary to perform the job of firefighter.²⁵ Since numerous candidates took both Written

February 21, 2008, p.170.

²⁵ Clearly, the City believed that everyone who passed Written Exam 7029 had the ability necessary to perform the job because the City exhausted the Exam 7029 eligibility list. While the City did not exhaust the Exam 2043 eligibility list, it treated anyone who passed both Written Exam 2043 and the PPT as eligible for hire (unless disqualified by some other step in the process, such as the background investigation). See Rule 30(b)(6) Dep. Test. of the City (T. Patitucci, désignee), p.145; Dep. Test. of Carol Wachter, January 17, 2008, pp.112-13.

-19-

Exams 7029 and 2043, one can address this issue by looking at the degree to which candidates who took both written examinations were consistently deemed either qualified (*i.e.*, passed both written examinations) or unqualified (*i.e.*, failed both) by the written examinations.

28. A total of 2,667 candidates took both Written Exams 7029 and 2043. Clearly, using a cutoff score of 84.705 for Written Exam 7029 and then using a cutoff score of 70 for Written Exam 2043 (which is purportedly the same test)²⁶ suggests that the 84.705 cutoff score is not meant to represent the minimum level of cognitive performance necessary to do the job of firefighter.²⁷ Among those who took both written examinations, 54.8 percent (17 out of 31) of those who failed Written Exam 2043 scored 70 or above on Written Exam 7029. Of those who scored below 70 on Written Exam 7029, 75.9 percent (44 out of 58) passed Written Exam 2043. Thus, of those failing either examination at a 70 percent cutoff score, 81.3 percent (61 out of 75) failed one written examination but passed the other. Since these figures reflect a single candidate taking different administrations of what is purportedly the same test, the high degree of inconsistency affirmatively highlights the fact that the cut-off scores on the written examinations do not reliably predict the presence or absence of the minimum cognitive skills and abilities to do the job.

29. As stated previously, the City has developed a new written examination, Written Exam 6019. One-hundred and fifty-six candidates, who scored below 70 on either Written Exam

²⁶ See Letter from G. Pestana to E. Yorke, dated January 9, 2004, stating that the two written examinations are "substantively identical."

²⁷ Of course, the City does not make such a claim. The data shows that 92.9 percent of those who scored below 84.705 on Written Exam 7029 and then took Written Exam 2043 passed Written Exam 2043, as would be expected since most of those who failed Written Exam 7029 scored above 70 on the examination.

-20-

7029 or Written Exam 2043 took Written Exam 6019. Of those 156 candidates, approximately 85 percent (actually, 84.6 percent) of those who previously scored below 70 on Written Exam 7029 or Written Exam 2043 passed Written Exam 6019 and are now considered by the City to be qualified²⁸ to be a firefighter for the FDNY.

ISSUE 4(a): THE PHYSICAL AND WRITTEN EXAMS WERE NOT GIVEN EQUAL WEIGHTS IN THE RANKING PROCESS, AND WERE NOT EVEN GIVEN THE SAME WEIGHTS FOR EXAMS 7029 AND 2043.

30. In creating the combined written examination and PPT score for the purpose of ranking candidates on the Exam 7029 and 2043 eligibility lists, the City did not assign equal weights to the PPT and the written examinations, as it intended. In fact, the PPT was given about twice the weight of Written Exam 7029. Conversely, Written Exam 2043 was given about 10 percent more weight than the PPT.²⁹

31. In combining the written examination and PPT score, the City standardized the scores and then averaged the standardized scores. *See* Appendices D and E for the City's explanation of this process for Exams 7029 and 2043, respectively. Correct standardization would have resulted in the written examination and the PPT being weighted equally. However, the City incorrectly used the written scores of all candidates who took the written examination and the PPT scores of all candidates who took the PPT to standardize the scores. *See* Appendices

²⁸ Of course, candidates must pass the other steps in the process, such as the background investigation, in order to be hired.

²⁹ Neither the weight actually given the written exam and PPT for Exam 7029 nor those actually given the written exam and PPT for Exam 2043 reflect the relative importance of cognitive and physical abilities to the firefighter job as determined by the 2005 DCAS survey referred to previously. *See* Firefighter Survey Requirements of Physical vs. Cognitive Abilities, Pl.'s Dep. Ex. 243.

-21-

C and D; *see also* Rule 30(b)(6) Dep. Test. of the City (T. Patitucci, designee), pp. 155-56. The problem is one of "mixing apples and oranges."³⁰ A candidate who failed the written examination could not take the PPT, and a candidate who failed the PPT was not on the eligibility list to be ranked. To properly weight each test for ranking purposes, the standardization should have been based on the common set of written examination and PPT passers who were on the eligibility list to be ranked. Because of this methodological error, the PPT and written examinations were not given equal weight. The actual weight given to each component for each of the examinations is delineated below.³¹ The lower weight for Written Exam 7029 is due to the higher cutoff score (84.705) used for Written Exam 7029.³²

³⁰ Using all the PPT scores is also a problem because the scores of those who failed were all truncated to 62.5, so a true score distribution did not exist. Dr. Cline also noted this difficulty in her deposition testimony. *See* Dep. Test. of Catherine Cline, February 21, 2008, pp. 67-69.

³¹ The effective weight of each component is the squared correlation (referred to as the coefficient of determination) of each component with the final weighted composite score (*i.e.*, the combination of the written and physical scores) scaled to sum to one.

³² If the standard deviation used is larger than the correct standard deviation, the effective weight of the test will be less than its planned weight (assuming the correct standard deviation is used to weight the other test(s)). Conversely, if the standard deviation used is smaller than the correct standard deviation, the effective weight of the test will be greater than its planned weight. The largest difference between the standard deviation used and the correct standard deviation occurred for Written Exam 7029. Because the City used such a high cutoff score on Written Exam 7029, the standard deviation used (the standard deviation of all written exam takers) was substantially larger than the correct standard deviation (the standard deviation of the group restricted to those on the eligibility list), and the effective weight of Written Exam 7029 was substantially less than the planned 50% weight.

-22-

Examination	Actual Weight in Creating Combined Scores for Ranking on Eligibility List	
	Written Exam	PPT Exam
7029	38%	62%
2043	55%	45%

32. Moreover, the majority of the candidates on the eligibility lists for both Exams 7029 and 2043 scored 100 on the PPT. For Exam 7029, 69.5 percent of the candidates on the eligibility list scored 100 on the PPT. On the Exam 2043 eligibility list, 75.3 percent of the candidates scored 100 on the PPT. Thus, setting aside bonus points, for the group of approximately 70 percent of the candidates on each eligibility list who scored the highest on the PPT, ranking was based solely on the candidates' scores on the written examination.³³

ISSUE 4(b): INCORRECTLY ANSWERING A SINGLE QUESTION ON WRITTEN EXAM 7029 OR 2043 OR FAILING A SINGLE PPT EVENT BY ONE SECOND HAD AN UNJUSTIFIABLY LARGE IMPACT UPON A CANDIDATE'S RANKING.

33. Table 5 displays a rank-order list for Exam 7029 based on the City's weighting of the written examination and the PPT, excluding all bonus points. Table 6 contains a similar list for Exam 2043. As Table 5 shows, 439 persons scored 100 on Written Exam 7029 and 100 on the PPT. If a candidate who scored 100 on Written Exam 7029 had failed one component of the PPT, their PPT score would have fallen to 87.5 and their maximum rank would have fallen from 1 to 3,603 on the list. Failing a second PPT component would have dropped them to 5,471 on the list. In this regard, it should be noted that performing one PPT event one second more slowly

³³ Thus, as the City apparently was aware, to a great extent the PPT was serving as a pass/fail test with a 100 being the passing score. See Meeting Notes on PPT, dated March 1, 2002, Pl.'s Dep. Ex. 537, p. 2.

-23-

could be the difference between a candidate passing all eight PPT events (and receiving a PPT score of 100), and failing one event (and receiving a PPT score of 87.5).

34. Similarly, a candidate who scored 100 on the PPT but got one question wrong on Written Exam 7029 would have fallen from a maximum rank of 1 to a maximum rank of 440. A second wrong answer on Written Exam 7029 (*i.e.*, a score of 98.82) would have dropped the candidate's rank to a maximum of 1,146; four wrong answers to Written Exam 7029 (*i.e.*, a score of 95.294) would have dropped the candidate's maximum rank to 2,454 on the list. Given the standard error of measurement of Written Exam 7029 (*i.e.*, 2.64),³⁴ a wrong answer to four questions on Written Exam 7029 is within the range of normal variation in scores due to chance (and may not represent any true difference between individuals).

35. The standard error of measurement of Written Exam 2043 is 2.63, which is almost identical to that of Written Exam 7029. As Table 6 shows, the scores for Written Exam 2043 were similarly clustered at the top, so four incorrect answers rather than a perfect score (which would constitute a drop from a score of 100 to a score of 95.294 on Written Exam 2043) with a 100 on the PPT, would have dropped a candidates' rank from a maximum rank of 1 to a maximum rank of 1,713. Similarly, a failure on a single component of the PPT rather than a perfect score on both the PPT and Written Exam 2043 would have dropped a candidate from a maximum rank of 1 to a maximum rank of 3,439; failure on a second component of the PPT would have dropped the candidate to a maximum rank of 5,776. Clearly, the ranking process resulted in large differences in rank due to small changes in performance on the written

³⁴ The standard error of measurement stated above was computed from candidates' answers to each test item based on the Kuder Richardson 20 formula.

-24-

examinations or the PPT.

36. To further examine the effect of small changes in examination scores on ranks, I studied the 2,667 test takers who took both Written Exam 7029 and 2043. The mean score of those who took both tests was 91.52 on Written Exam 7029, with a standard deviation of 7.52. While their average score on Written Exam 7029 was almost identical to the average of all test takers (*i.e.*, 91.18), their standard deviation was statistically significantly smaller (7.52 versus 9.46).³⁵

37. Even though the two tests purportedly measure the same cognitive abilities and are constructed in the same way,³⁶ the rankings of the candidates based on one exam are very different from the rankings of the same candidates based on the other exam. In the extreme, when comparing a candidate's rank on Written Exam 7029 to his rank on Written Exam 2043, we find that 12 candidates dropped more than 2,000 places in rank (out of 2,667 total places) and nine candidates rose more than 2,000 places in rank. Although not as extreme, the typical (*i.e.*, median) change in ranking is still large. Comparing the rankings of all of the 2,667 candidates who took both examinations, we find the typical change in an individual's rank between Written

³⁵ This makes sense. Candidates who scored very high on Written Exam 7029 were more likely to be hired and, therefore, not to take Written Exam 2043. Candidates who scored very low on Written Exam 7029 may have been discouraged from taking Written Exam 2043.

³⁶ The correlation between the two test scores is an indicator of the reliability of the written examinations' design and construction, since the two examinations were supposed to measure the same cognitive abilities. It should be noted, however, that the correlation is not a perfect indicator of reliability, since there was such a long time lapse between the administrations of the two exams. Since the standard deviation of the multiple test takers was restricted, I corrected the correlation for restriction in range in computing the correlation between the Written Exams 7029 and 2043 scores. The corrected correlation between the two written examination scores for the 2,667 individuals who took both written examinations was 0.646.

-25-

Exam 7029 and Written Exam 2043 was 458 positions (a change of about 20 percent up or down) on the list. Apparently, large changes in rank occurred primarily due to chance.³⁷

³⁷ Some of this change could be due to the time lapse between the administrations of the two examinations and the possibility that, as a result, the candidates' skills and knowledge might have actually changed between the administrations of the examinations.

TABLE 1

ITEM INTERCORRELATIONS WITHIN AND BETWEEN COGNITIVE ABILITY DOMAINS

EXAM 7029

	DEDUCTIVE REASONING	INDUCTIVE REASONING	INFORMATION ORDERING	MEMORIZATION	PROBLEM SENSITIVITY	SPATIAL ORIENTATION	VISUALIZATION	WRITTEN COMPREHENSION	WRITTEN EXPRESSION
DEDUCTIVE REASONING	0.139								
INDUCTIVE REASONING	0.136	0.156							
INFORMATION ORDERING	0.135	0.133	0.119						
MEMORIZATION	0.078	0.072	0.071	0.082					
PROBLEM SENSITIVITY	0.125	0.124	0.114	0.073	0.109				
SPATIAL ORIENTATION	0.147	0.145	0.138	0.076	0.128	0.163			
VISUALIZATION	0.129	0.132	0.125	0.071	0.114	0.139	0.120		
WRITTEN COMPREHENSION	0.148	0.141	0.137	0.078	0.127	0.153	0.136	0.144	
WRITTEN EXPRESSION	0.153	0.157	0.145	0.088	0.136	0.157	0.146	0.155	0.163

TABLE 2

ITEM INTERCORRELATIONS WITHIN AND BETWEEN COGNITIVE ABILITY DOMAINS

EXAM 2043

	DEDUCTIVE REASONING	INDUCTIVE REASONING	INFORMATION ORDERING	MEMORIZATION	PROBLEM SENSITIVITY	SPATIAL ORIENTATION	VISUALIZATION	WRITTEN COMPREHENSION	WRITTEN EXPRESSION
DEDUCTIVE REASONING	0.158								
INDUCTIVE REASONING	0.150	0.222							
INFORMATION ORDERING	0.169	0.215	0.259						
MEMORIZATION	0.075	0.060	0.073	0.087					
PROBLEM SENSITIVITY	0.137	0.189	0.192	0.061	0.184				
SPATIAL ORIENTATION	0.128	0.108	0.126	0.078	0.102	0.172			
VISUALIZATION	0.128	0.116	0.134	0.080	0.117	0.124	0.219		
WRITTEN COMPREHENSION	0.142	0.143	0.154	0.089	0.128	0.114	0.116	0.140	
WRITTEN EXPRESSION	0.124	0.116	0.133	0.067	0.106	0.121	0.102	0.113	0.117

EXHIBIT W

1

1

2 IN THE UNITED STATES DISTRICT COURT
3 FOR THE EASTERN DISTRICT OF NEW YORK
4 Civil Action No. 07-CV-2067

-----x
4 UNITED STATES OF AMERICA,

5 PLAINTIFF,

6 AND

7 VULCAN SOCIETY, INC., FOR ITSELF AND ON
8 BEHALF OF ITS MEMBERS, CANDIDO NUNEZ,
9 MARCUS HAYWOOD AND ON BEHALF OF A CLASS
10 OF ALL OTHERS SIMILARLY SITUATED,

11 PLAINTIFF-INTERVENORS,

12 V.

13 CITY OF NEW YORK, FIRE DEPARTMENT OF THE
14 CITY OF NEW YORK, NEW YORK CITY
15 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
16 SERVICES, MAYOR MICHAEL BLOOMBERG AND NEW
17 YORK CITY FIRE COMMISSIONER NICHOLAS
18 SCOPPETTA, IN THEIR INDIVIDUAL AND
19 OFFICIAL CAPACITIES,

20 DEFENDANTS.

21 -----x
22 January 17, 2008
23 9:47 a.m.

24

25 DEPOSITION of CAROL WACHTER, taken by
Plaintiff, pursuant to Rule 30(b)(6)
Notice, held at the United States
Attorney for the Eastern District of New
York, 271 Cadman Plaza East, Brooklyn,
New York, before Jamie Ann Stanton, a
Shorthand Reporter and Notary Public of
the State of New York

26 * * *

27

28

1 C. Wachter

2 Q (Hanging.)

3 A (Reviewing exhibit.)

4 Q Have you seen this document
5 before?

6 A Well, my signature is on it, so
7 I assume I did.

8 Q What is it?

9 A "Summary of Written Test Scores
10 for Exam 7029 Firefighter."

11 Q I would like to call your
12 attention to the top of the exhibit. It
13 states that the pass mark for the written
14 part of Exam 7029 is 84.705; is that
15 correct?

16 A I guess it is.

17 Q What was your involvement, if
18 any, in the decision to set the pass mark
19 for Written Exam 7029 at 84.705?

20 A Okay. From my recollection, I
21 believe we had discussions with the Fire
22 Department as to how to set this pass
23 mark. The Fire Department gave us their
24 estimate of how many hires they expected
25 to make over the course of a four-year

1 C. Wachter

2 list. Using past experience, we worked

3 backwards from that number to determine

4 how many people we would need to pass the

5 multiple choice test in order to come up

6 with a list of that number. Taking into

7 consideration the percentage of people who

8 would likely show up for the physical, the

9 percentage that would pass the physical,

10 the percentage of those people who would

11 meet the requirements, actually show up

12 for their interviews, pass investigation

13 and the medical and therefore be ready to

14 be eligible to be appointed.

15 Q It sounds to me like there were

16 some sort of calculations involved in

17 order to determine --

18 A Right.

19 Q -- make this determination?

20 A Right.

21 Q Were those calculations

22 committed to paper?

23 A I couldn't -- I really don't

24 remember at this point.

25 Q Now, you mentioned we had

1 C. Wachter

2 regarding whether it was appropriate to
3 limit the number of candidates based on
4 the written test score alone?

5 A I don't remember.

6 Q Did you ever express an opinion
7 about whether it was appropriate to limit
8 the number of candidates based on the
9 written test score alone?

10 A Did I ever express an opinion?

11 Q Yes, ma'am.

12 A I don't -- I really don't
13 remember.

14 Q Were any validity studies
15 conducted with respect to the pass mark
16 for Written Exam 7029?

17 A No.

18 Q Was any job analysis done in
19 connection with the setting of the pass
20 mark for Written Exam 7029?

21 A Not that I'm aware of.

22 Q Setting the pass mark of 84.705,
23 did DCAS consider whether that cutoff
24 corresponded to the minimum level of
25 skills measured by the written exam that

1 C. Wachter

2 are necessary to perform the job of
3 entry-level firefighter successfully?

4 A I don't know that we did that
5 specifically. You know, our ordinary pass
6 marks would have been 70. So if one would
7 assume that 70 was the minimal competency
8 level, then certainly 84 would have not
9 been less than the competency level.

10 Q Is it fair to say that 84.705
11 was in excess of the minimum competency
12 level?

13 A I think you could say that.

14 Q Did anyone at the Fire
15 Department or within DCAS express concerns
16 that the setting of the pass mark at
17 84.705 might increase the likelihood of
18 litigation regarding Exam 7029?

19 A I don't remember that being an
20 issue.

21 Q Now, why did you want to limit
22 the number of candidates for the physical
23 exam to 15,000?

24 A There were probably a couple of
25 issues involved that I can think of

1 C. Wachter

2 to that particular time.

3 Q Let me ask it this way: In this

4 instance, with respect to Written Exam

5 7029, do you think it would have been

6 responsible to go with a lower pass mark

7 for the written examination --

8 A I think we could have --

9 Q I'm sorry.

10 A I thought you were finished.

11 Q Would it have been possible to

12 go with the lower pass mark for the

13 written examination, and with this lower

14 pass mark, still select only candidates

15 who are qualified for the position at

16 least on the basis of the written exam?

17 A Yes.

18 Q So to be more clear: You

19 believe that by setting the pass mark at

20 84.705, qualified candidates were excluded

21 from consideration for processing?

22 A Well, they probably would not

23 have been reached for appointment any way,

24 so it's kind of moot.

25 Q Was the eligibility list for

180

1 C. Wachter

2 Exam 7029 exhausted?

3 A I believe it was, but I'm not a

4 hundred percent sure. After 9/11 the Fire

5 Department had a need for more hiring than

6 they had anticipated, so I would think

7 that it was exhausted.

8 Q So given that the eligibility

9 list for Exam 7029 was exhausted, would it

10 be fair to conclude that setting the pass

11 mark at 84.705 eliminated qualified

12 candidates for the position who might

13 otherwise have been appointed but for

14 setting the pass mark at 84.705?

15 A Isn't hindsight wonderful?

16 Q Yes, ma'am. Is that an unfair

17 conclusion?

18 A No.

19 MR. REESE: I don't have any

20 further questions.

21 EXAMINATION BY

22 MR. EDWARD:

23 Q If I can direct your attention,

24 Ms. Wachter, to Plaintiff's Exhibit 25.

25 Do you see there, at the top, it

181

1 C. Wachter

2 says "Summary of Written Test Scores"?

3 A Yes.

4 Q It says "84.705"?

5 A Correct.

6 Q I want to know the reason why it

7 was set at 84.705. I think we covered

8 this. I want to read for you the bottom

9 paragraph down there.

10 It says: "Reasons for

11 recommendation. The Fire Department has

12 requested that a total of 15,000

13 candidates from both test and the

14 promotion test to be called to the

15 physical. At this pass mark, the Fire

16 Department's request is met."

17 Is that the reason why the pass

18 mark was set at 84.705?

19

20 [Continued on the next page to

21 allow for signature line and jurat.]

22

23

24

25

182

1

2

3 A Yes.

4 MR. EDWARD: No further

5 questions.

6

7 [TIME NOTED: 2:28 p.m.]

8

9 _____

10 CAROL WACHTER

11

12

13 _____
Subscribed and sworn to
before me this _____
14 day of _____,
2008.

15

16 _____
Notary Public

17

18

19

20

21

22

23

24

25

EXHIBIT X

1

2 IN THE UNITED STATES DISTRICT COURT
3 FOR THE EASTERN DISTRICT OF NEW YORK
4 Civil Action No. 07-CV-2067

-----x
4 UNITED STATES OF AMERICA,

5 PLAINTIFF,

6 AND

7 VULCAN SOCIETY, INC., FOR ITSELF AND ON
8 BEHALF OF ITS MEMBERS, CANDIDO NUNEZ,
9 MARCUS HAYWOOD AND ON BEHALF OF A CLASS
10 OF ALL OTHERS SIMILARLY SITUATED,

11 PLAINTIFF-INTERVENORS,

12 V.

13

14 CITY OF NEW YORK, FIRE DEPARTMENT OF THE
15 CITY OF NEW YORK, NEW YORK CITY
16 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
17 SERVICES, MAYOR MICHAEL BLOOMBERG AND NEW
18 YORK CITY FIRE COMMISSIONER NICHOLAS
19 SCOPPETTA, IN THEIR INDIVIDUAL AND
20 OFFICIAL CAPACITIES,

21 DEFENDANTS.

22 -----x
23 January 14, 2008
24 9:00 a.m.

25

19 DEPOSITION of MATTHEW MORRONGIELLO,
20 taken by Plaintiff, pursuant to Rule
21 30(b)(6) Notice, held at the United
22 States Attorney for the Eastern District
23 of New York, 271 Cadman Plaza East,
24 Brooklyn, New York, before Jamie Ann
25 Stanton, a Shorthand Reporter and Notary
Public of the State of New York

* * *

1 M. Morrongiello

2 Supervisor, Sanitation. Those are the
3 ones that come to mind.

4 Q So it sounds like you had
5 developed a written exam for entry-level
6 sanitation worker; is that correct?

7 A Correct.

8 Q Is that the only entry-level
9 exam you had developed before 7029?

10 A That's all I can recall right
11 now.

12 Q And before when you said you had
13 done maybe four or five job analyses
14 already at the time you were working on
15 7029, were those for entry-level positions
16 or promotional positions?

17 A That would be a combination.

18 Q How many for entry-level?

19 A Right now, sanitation worker
20 comes to mind. That's all I can recall.

21 Q Do you know what a
22 criterion-related validity study is?

23 A Yes.

24 Q Have you ever conducted a
25 criterion-related validity study?

1 M. Morrongiello

2 A Not that I'm aware of.

3 Q Do you know what a construct
4 validity study is?

5 A Yes.

6 Q Have you ever conducted a
7 construct validity study?

8 A Not that I'm aware of.

9 Q Have you ever conducted a
10 content validity study?

11 A Not that I'm aware of.

12 Q I want to talk a little bit
13 about your training for various things
14 like job analysis and test development.
15 What I mean by "training," as opposed to
16 formal education, is something that's not
17 a course at a college or university. It
18 could be on-the-job training or any other
19 training outside of a university or
20 college.

21 My first question is: What
22 training have you had in conducting job
23 analyses?

24 A Well, we really consider it
25 on-the-job training. The training I had

1 M. Morrongiello

2 A No.

3 Q Now, maybe this is obvious, but

4 I just want to make sure we are clear. On

5 the document, on Exhibit 54, it says "Test

6 Development Report." I just want to make

7 sure that we characterize the report in

8 the proper way.

9 I take it you consider this a

10 test development report, correct?

11 A Correct.

12 Q Was it intended to be anything

13 other than a factual narrative,

14 essentially, of what you did in order to

15 develop or prepare the written exam used

16 for Exam 7029?

17 A Intended to be anything -- no.

18 Q Was it intended in any way to be

19 a validity report?

20 A No.

21 Q To your knowledge, is Exhibit 54

22 the only written report ever prepared

23 regarding Exam 7029?

24 A To the best of my knowledge, it

25 is.

1 M. Morrongiello

2 expensive to do.

3 My current question is whether

4 you have ever heard that discussed by

5 anyone?

6 A I don't recall hearing that

7 discussed.

8 Q Have you ever done any analysis

9 of whether a score of 84.705 on Written

10 Exam 7029 corresponded to the minimum

11 level of the tested abilities necessary to

12 perform the FDNY firefighter job

13 successfully?

14 A Repeat that once more.

15 Q Sure. Have you ever done any

16 analysis of whether the cutoff score of

17 84.705 used on Written Exam 7029

18 corresponds to the level of the tested

19 abilities necessary to perform the

20 firefighter job successfully?

21 A No.

22 Q Have you ever done any analysis

23 of whether a score of 70 on written Exam

24 2043 corresponds to the minimum level of

25 the tested abilities necessary to perform

1 M. Morrongiello

2 the firefighter job successfully?

3 A No.

4 Q Do you know whether anyone has

5 ever attempted a criterion-related

6 validity study with respect to Exam 7029?

7 A I don't know.

8 Q Do you know whether anyone who

9 has ever attempted a criterion-related

10 validity study regarding Exam 2043?

11 A I don't know.

12 Q Do you know whether anyone has

13 ever attempted a construct validity study

14 regarding either Exam 7029 or Exam 2043?

15 A I don't know.

16 Q We've been told that for

17 purposes of ranking on the Exam 7029

18 eligibility list, the written exam and the

19 physical exam, the PPT, were given equal

20 weight; is that correct?

21 A Correct.

22 Q Why were they given equal

23 weight?

24 A Well, my understanding is

25 that -- just give me one second. My

1 M. Morrongiello

2 that we were going to use that physical
3 for 7029.

4 Q Did someone tell you that, or is
5 that some conclusion that you came to,
6 that the physical was still appropriate?

7 A No, I was told that.

8 Q Who told you that?

9 A I don't remember who said that
10 specifically.

11 Q After that note, there is
12 something that is kind of off to the right
13 and circled. It appears to say: "First
14 day scale" -- can you read to me what is
15 in that circled portion?

16 A You mean starting at the left?

17 Q Yes.

18 A "First day scale for tasks. And
19 then there is an arrow. Then reading to
20 the right, it says: "Just performance."
21 There is a "2" circled above it. And
22 "frequency," there is a "3" circled above
23 it. And then the word "importance."

24 Q In your job analysis
25 Questionnaire, you did not include a first

1 M. Morrongiello

2 analysis Questionnaire.

3 And the question?

4 Q When you referred to "all the

5 Fleishman's," you are only referring to

6 those twenty-one and not to the physical

7 abilities and the psychomotor abilities

8 that are also included in Dr. Fleishman's

9 taxonomy; is that correct?

10 A Correct.

11 Q Did you consider including any

12 noncognitive abilities or characteristics

13 in your list?

14 A I believe it was decided to go

15 with our -- what I'll call the standard

16 Fleishman ability list that we used for

17 the Open Competitive.

18 Q My question was: Did you

19 consider using anything else, though?

20 A I'm not sure.

21 Q Have you ever discussed with

22 anyone from the FDNY, the Fire Department,

23 the relative importance to the firefighter

24 job of noncognitive or personality

25 characteristics as compared to cognitive

130

1 M. Morrongiello

2 abilities?

3 A As part of the job analysis? Is

4 that what you mean?

5 Q No. Ever. Have you ever

6 discussed it with anyone from the FDNY?

7 A Just repeat that once more.

8 Q Sure. Did you ever discuss,

9 ever --

10 A Ever.

11 Q -- discuss with anyone from the

12 FDNY the relative importance of

13 noncognitive or personality

14 characteristics as compared to cognitive

15 abilities for the firefighter job?

16 A I don't recall having those

17 discussions.

18 Q Have you ever discussed that

19 with anyone else?

20 A Well, in this case, we did talk

21 about the possibility of doing -- that we

22 want to do a job analysis for the

23 physical. That's one example. That's all

24 that come to mind.

25 Q So you don't recall ever talking

1 M. Morrongiello

2 Q Now, the firefighters who were
3 given the job analysis Questionnaire were
4 asked to rate the importance of the
5 abilities -- I think the term I would use
6 is globally, importance to the entire job
7 rather than to any particular task or task
8 category, correct?

9 A The scale is asking them to
10 compare it to the other abilities, so can
11 you repeat what you just said?

12 Q Right. My question was that
13 they were asked to rate the importance, to
14 use this scale, to rate the importance of
15 the abilities to the performance of their
16 whole job, not to the performance of any
17 particular task or task category; is that
18 correct?

19 A That's correct.

20 Q Were the firefighters given any
21 guidance on how to judge the importance of
22 each ability to the job as a whole?

23 A I don't recall saying anything
24 specific about the importance scale.

25 Q So is it true, then, that they

1 M. Morrongiello

2 weren't given a deposition of what
3 somewhat important meant in the rating
4 scale?

5 A That's true. I'm not aware of a
6 definition that we would have given.

7 Q Or they weren't given an
8 explanation of the difference between a
9 rating of important to the performance of
10 the job versus critical to the performance
11 of the job; is that right?

12 A I'm not aware of any instruction
13 that we would have given for that.

14 Q Were the firefighters who filled
15 out the job analysis Questionnaire given
16 any examples of what those ratings would
17 mean?

18 A For this administration, I don't
19 recall being that specific, talking about
20 the questionnaire.

21 Q Essentially the same question
22 for the importance scale for tasks, which
23 was on the previous page, Scale B.

24 Were they given any definitions
25 or explanations or examples of what the

1 M. Morrongiello

2 various points on the importance scale for
3 tasks meant?

4 A I don't recall giving any
5 instruction for that.

6 Q Now, if we go back into the body
7 of your report on page 10.

8 A (Complying.)

9 Q It looks like the scales that we
10 saw in the actual job analysis
11 Questionnaire, the A, B, C, D scales were
12 converted to numeric scales of 1 through
13 4.

14 Is that what you did?

15 A I just want to make sure. Yes,
16 they were, right, converted to numeric.

17 Q Why did you convert the
18 alphabetic scales to numeric?

19 A Well, that allows us to get a
20 number reading on the scales -- on the
21 responses.

22 Q So that you can do things like
23 averaging them; is that correct?

24 A Correct. Correct.

25 Q So who chose which numeric value

1 M. Morrongiello

2 My question is: Was the Linking

3 Panel given any definition or explanation

4 of what "somewhat important," "important"

5 and "critical" mean?

6 A Not that I am aware of.

7 Q Were they given any examples of

8 what those mean?

9 A I'm not aware of any.

10 Q So other than giving the

11 definitions that we see here on the page

12 for the ratings scale, was anything done

13 to ensure that the members of the Linking

14 Panel understood how to rate "somewhat

15 important" versus "important" versus

16 "critical"?

17 A Can you repeat that?

18 Q Yes. I think it got garbled in

19 the middle somewhere, but the question

20 essentially was: Other than just what is

21 here in writing as the definitions of 0

22 through 3 on the ratings scale, was the

23 Linking Panel given any explanation of how

24 to distinguish between "critical,"

25 "important" and "somewhat important" to

252

1

2 rate the abilities?

3 A I don't remember giving any.

4 Q Now, if we can go to the page

5 Bates numbered C-002546, which is where

6 the ability list begins --

7 MS. SCOLNICK: Sharon, would

8 this be a good place to take a break?

9 MS. SEELEY: Yes. I guess we

10 will stop there and continue some

11 other day.

12 [TIME NOTED: 4:10 p.m.]

13

14

15 MATTHEW MORRONGIELLO

16

17

18 Subscribed and sworn to
19 before me this _____
day of _____,
20 2008.

21

22 Notary Public

23

24

25

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK
Civil Action No. 07-CV-2067

-----x
THE UNITED STATES OF AMERICA,
Plaintiff,

and
VULCAN SOCIETY, INC., for itself and on
behalf of its members, CANDIDO NUNEZ,
MARCUS HAYWOOD and on behalf of a Class
of All Others Similarly Situated,
Plaintiff-Intervenors

-against-

CITY OF NEW YORK, FIRE DEPARTMENT OF THE
CITY OF NEW YORK, NEW YORK CITY
DEPARTMENT OF CITYWIDE ADMINISTRATIVE
SERVICES, MAYOR MICHAEL BLOOMBERG and
NEW YORK CITY FIRE COMMISSIONER NICHOLAS
SCOPETTA, in their Individual and
Official capacities,
Defendants.

-----x
January 29, 2008
10:42 a.m.

Continued Deposition of MATTHEW

MORRONGIELLO, taken by the Plaintiff, pursuant
to Rule 30(b)(6) Notice, at the offices of the
United States Attorney for the Eastern
District of New York, 271 Cadman Plaza East,
New York, New York, before David Levy, CSR, a
Notary Public of the State of New York.

1 Morrongiello

2 A. Yes.

3 Q. Do you recall that Linking panel?

4 A. That used the same list?

5 Q. Yes.

6 A. I can give you the names of open
7 competitive exams where in all likelihood, we
8 used the same list. One that comes to mind is
9 the police officer exam. Right now, that's all I
10 can think of.

11 Q. And do you recall that particular
12 Linking panel session?

13 A. No.

14 Q. After the Linking panel session for
15 exam 7029 was over, did you conduct any
16 statistical analyses of the reliability or
17 interrater agreement among the ratings of the
18 twelve Linking panel members?

19 A. No.

20 Q. Did you look at all at the extent to
21 which the ratings of the Linking panel members
22 agreed or did not agree with each other?

23 A. Um -- I don't recall doing that.

24 Q. Okay. I'm going to ask you some
25 questions now about the exhibits that we've

Morrongiello

offer is that perhaps in either, in the training that firefighters receive, there could be some written material that states how to trim broken glass, the procedures to use, which tools to use. That's an explanation that I can offer.

Q. So let me ask this, then:

Are you saying that a Linking panel member might have rated written comprehension as important to some degree to any given task if they read some written material about that task in training?

A. Possibly. You'd have to ask the people who filled it out. But --

Q. Is that what they were supposed to do?

A. Well, based on the explanation or the definition of the -- of the ability, if that involved any material that they would need to read, then they most likely would apply that.

Q. So the task, then, doesn't just involve performing the task itself. It also involves training for that task, is that correct?

A. I'm just offering an explanation how a, someone doing the job could rate something

1 Morrongiello

2 like written comprehension critical for a task
3 that doesn't just come out and state --

4 Q. And what I'm trying to get at, is, is
5 that what you intended? Would that be an
6 appropriate rating? Was a task supposed to refer
7 only to performance of the task rather than
8 training for the task or writing about the task
9 later?

10 A. Many times when we give directions,
11 we ask people to think about the job they are
12 currently doing. I guess the best explanation I
13 could give you and my best response would be, if
14 they feel that, in their job they are currently
15 doing, if this ability is important for that
16 particular task, then they would rate it
17 appropriately.

18 Q. But what I intended my question to
19 be, at least, maybe I didn't ask it clearly, was
20 whether that is what they are supposed to be
21 doing, what you want them to be doing; consider
22 only the task in the performance of the task, or
23 also consider things that are like training for
24 the task or writing later about the task, but not
25 the actual performance of the task.

Page 317

1 Morrongiello

2 A. We generally would like them to just
3 focus in on performing that particular task.
4 That's what we ask them to respond to. So that's
5 really where we're going. We generally say,
6 "Right now on your job, to perform this task,
7 would you would you rate these abilities?"

8 Q. I think I understand. But you don't
9 know what was actually going on in the mind of
10 any one of these individuals when they did the
11 ratings, correct?

12 A. That's correct. That's correct.
13 Could I just add something to that? Generally,
14 that's why we try to have groups involved. When
15 we do surveys, sometimes having a number of
16 people, you may get different responses.

17 Q. Okay. Let me ask you, still looking
18 at Exhibit 85, when you got the ratings from
19 Rater 9, did you notice that this rater gave a
20 pretty much large majority of the ratings as
21 rating 3, "Critical"?

22 A. Probably not.

23 Q. Did you actually review the ratings
24 given by the Linking panel members to see if they
25 made sense?

Page 318

Morrongiello

A. I generally will look through the packets. Generally, though, not to see if they make sense.

Q. For what purpose would you look through them?

A. There could have been some blanks left. That's the sort of thing we would add, return it to the -- to the panel member and point it out and --

Q. So you would actually review it when they turned it in to you at the session.

A. Generally. Generally we try to, and for the reason I just stated. If they leave us and there are some blanks, we, you know, it could be hard to get them back or --

Q. Is there anything else that you review the packets for?

A. I just recall, like I said, looking over to make sure that things are complete.

Q. Just to make sure I'm clear on this, did you talk to any of the Linking panel members about any of their ratings?

A. I don't recall.

Q. Is that something you would normally

1 Morrongiello

2 Q. Well, the item writers are not
3 specialists in knowing what reading levels are.

4 A. No, that's true. Like, for example,
5 we're not asking them to keep the test items
6 appropriate to what someone might need on the
7 job. Maybe that's a better way to describe it.
8 So we don't ask them to say, "Could a tenth
9 grader read this." It's more towards what would
10 people need to know for the job.

11 Q. And how about the examiners, what
12 would the examiners do to make sure it's at the
13 right level?

14 A. It's probably similar. I mean, I
15 think we do have, you know, examiners, I think
16 the safe to say, have at least an idea of
17 something that might be high school level or
18 something that maybe someone with a Bachelor's
19 degree might, if you need a Bachelor's degree to
20 read something.

21 So there's nothing formal about it.
22 That's why we have a number of people review our
23 items to give their best judgement.

24 Q. How do they get that idea?

25 A. I'm sorry?

Page 457

1 Morrongiello

2 Q. The examiners, how do they get that
3 idea of what would be an appropriate level?

4 A. I guess it's based on experience.

5 Q. So it's sort of -- it's informal,
6 eyeballing-it level?

7 A. Yes. Except for, as I said, there
8 have been cases, I don't know if -- I talked
9 about that readability index. Could have done
10 that for other exams. I'm just not aware right
11 now if we did that for other exams, that more
12 formal readability index that we're talking
13 about.

14 Q. And as far as the 7029, I'm sorry,
15 tell me whether you did it in the 7029.

16 A. I may have. I'd have to look back at
17 the records.

18 Q. If you did, would you have expected
19 something like that to be included in your test
20 in Exhibit 54, test development report?

21 A. Unlikely, but it would be, if it was
22 done, it would be in the records. Or it should
23 be in the files.

24 Q. In preparing for 7029, did you review
25 0084?

Morrongiello

A. What part of 0084?

Q. Any part of 0084.

A. The test, the --

Q. The test.

A. The test. I can't recall specifically reviewing it. I -- I'm looking back, I -- I probably did, if not all the test, then some portions of it. But I just can't recall if I read through every page and every item.

Q. Was there an item analysis prepared for 0084 similar to Exhibit 47 for 7029?

MR. SAMPLE: Objection, speculation.

A. Most likely.

Q. Most likely yes?

A. Most likely yes. There would be an item analysis. That's standard for really all our tests, is to produce item analyses in most cases.

Q. Going back, how long did you have item analyses?

MR. SAMPLE: Objection, ambiguous.

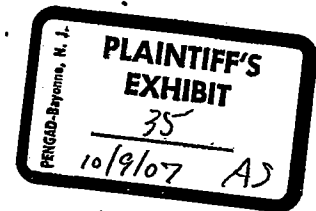
A. I can't recall. Since I've worked there, I seem to recall item analysis always

EXHIBIT Y

**FIREFIGHTER
EXAM. NO. 7029
EXPLANATION OF TEST SCORES**

Congratulations, you passed the examination for Firefighter, Exam. No. 7029. You had to achieve a written test score of 84.705 or greater in order to be eligible to take the physical test part of this examination. In order to pass the physical test, you had to pass 6 out of the 8 physical test events (75%). You had to pass both the written and physical tests in order to pass the overall examination. The enclosed Notice of Result card includes the following test information:

Final Average (Part H)
Written Test Score (Part E)
Physical Test Score (Part U)
Residency Credit, if applicable (Part R)
Legacy Credit, if applicable (Part Y)
Veterans Preference, if applicable
Adjusted Final Average
List Number



I. COMPUTING YOUR FINAL AVERAGE (Part E)

A. Computing your Standardized Written Test Score

Your standardized written test score was computed by subtracting the average of all candidates' written test scores (91.230) from your written test score (Part E), and dividing the result by the standard deviation of the written test (9.308). Round the product to 5 decimal places. Note that your standardized written test score may be either a positive or a negative number.

B. Computing your Standardized Physical Test Score

Your standardized physical test score was computed by subtracting the average of all candidates' physical test scores (87.262) from your physical test score (Part U), and dividing the result by the standard deviation of the physical test (15.582). Round to 5 decimal places. Note that your standardized physical test score may be either a positive or a negative number.

C. Computing your Combined Weighted Standard Score

1. Your Standardized Written Test Score was multiplied by the weight of the written test (.50) to produce your Weighted Standardized Written Test Score. Round to 5 decimal places.
2. Your Standardized Physical Test Score was multiplied by the weight of the physical test (.50) to produce your Weighted Standardized Physical Test Score. Round to 5 decimal places.
3. Your Weighted Standardized Written Test Score and your Weighted Standardized Physical Test Score were then added together to produce your Combined Weighted Standard Score. Round to 3 decimal places.

D. Computing your Transformed Score

Your Transformed Score was computed by multiplying your Combined Weighted Standard Score by 18.472906403940886699 and rounding the product to 3 decimal places. The product was then added to 83.74384236453 and rounded to 3 decimal places to produce your Transformed Score.

E. Computing your Final Average (Part H)

Your Final Average (Part H) was computed by adding your Residency Credit, if applicable (5 points) and Legacy Credit, if applicable (5 points) to your Transformed Score.

II. COMPUTING YOUR ADJUSTED FINAL AVERAGE

Your Adjusted Final Average was computed by adding your Veterans Preference, if applicable (5 points for Veterans; 10 points for Disabled Veterans) to your Final Average.

B-000007

EXAMPLE OF HOW TO COMPUTE YOUR ADJUSTED FINAL AVERAGE

Written Test Score (Part E) (50%)	Physical Test Score (Part U) (50%)	Residency Credit (Part R)	Legacy Credit (Part Y)	Veterans Preference V
96.750	100.000	5.000	5.000	

I. COMPUTING YOUR FINAL AVERAGE (Part H)

A. Computing your Standardized Written Test Score

$$\text{Standardized Written Test Score} = \frac{\text{Written Test Score (Part E)} - 91.230}{9.308}$$

$$\text{Standardized Written Test Score} = \frac{96.750 - 91.230}{9.308} = 0.59304 \text{ (Round to 5 decimal places)}$$

B. Computing your Standardized Physical Test Score

$$\text{Standardized Physical Test Score} = \frac{\text{Physical Test Score (Part U)} - 87.262}{15.582}$$

$$\text{Standardized Physical Test Score} = \frac{100.000 - 87.262}{15.582} = 0.81748 \text{ (Round to 5 decimal places)}$$

C. Computing your Combined Weighted Standard Score

$$1. \text{ Weighted Standardized Written Test Score} = \text{Standardized Written Test Score} \times .50$$

$$\text{Weighted Standardized Written Test Score} = 0.59304 \times .50 = 0.29652 \text{ (Round to 5 decimal places)}$$

$$2. \text{ Weighted Standardized Physical Test Score} = \text{Standardized Physical Test Score} \times .50$$

$$\text{Weighted Standardized Physical Test Score} = 0.81748 \times .50 = 0.40874 \text{ (Round to 5 decimal places)}$$

$$3. \text{ Combined Weighted Standard Score} = \text{Weighted Standardized Written Test Score} + \text{Weighted Standardized Physical Test Score}$$

$$\text{Combined Weighted Standard Score} = 0.29652 + 0.40874 = 0.705 \text{ (Round to 3 decimal places)}$$

D. Computing your Transformed Score

$$\text{Transformed Score} = (\text{Combined Weighted Standard Score} \times 18.472906403940886699) \text{ (Round to 3 decimal places)} + 83.74384236453$$

$$\text{Transformed Score} = (0.705 \times 18.472906403940886699) \text{ (Round to 3 decimal places)} + 83.74384236453 = 96.767 \text{ (Round to 3 decimal places)}$$

D. Computing your Final Average (Part H)

$$\text{Final Average} = \text{Transformed Score} + \text{Residency Credit, if applicable} + \text{Legacy Credit, if applicable}$$

$$\text{Final Average} = 96.767 + 5.000 + 5.000 = 106.767$$

II. COMPUTING YOUR ADJUSTED FINAL AVERAGE

$$\text{Adjusted Final Average} = \text{Final Average} + \text{Veterans Preference points, if applicable}$$

$$\text{Adjusted Final Average} = 106.767 + 5 = 111.767$$

B-000008

EXHIBIT Z

to increasing validity and to lowering adverse impact against black test takers. Some of these characteristics or abilities are: teamwork, responsibility, desire to learn, courage, resistance to stress, getting along with others, honesty, and medical interest. Dr. Schemmer also testified that "certainly tests of those constructs or categories were available during that time frame [for Exam 7029]."⁶⁷

Part V: The DCAS Job Analysis Does Not Support Validity

DCAS conducted only one job analysis for both Exams 7029 and 2043, and that one job analysis was seriously flawed in design, construction, and administration, to the point that the data summaries cannot be trusted and are not suitable to be used to support the content validity of the WT.

1. The Job Analysis Omitted Important Testable Abilities and Characteristics

The abilities rated by the Firefighters in the job analysis (and tested on the WTs) were chosen without any consultation with Firefighters and were limited to cognitive abilities (and specifically to the Fleishman cognitive abilities),⁶⁸ omitting such non-cognitive characteristics as dependability, integrity, or the ability to get along with others. Numerous studies done in other jurisdictions nationwide point to abilities DCAS did not even consider for inclusion on the WT. The 7029 job analysis process did not allow Firefighters to add to the abilities to be rated, and only rated abilities were included in the Test Plan (the outline of the number of questions testing for each ability).⁶⁹ There is no indication of the reason for the omission of non-cognitive abilities in the Test Development Report for Written Test 7029.⁷⁰ Omitting non-cognitive abilities from the 7029 job analysis precluded their being included in DCAS exams 7029 and 2043,⁷¹ and assured that the DCAS Written Tests would not measure non-cognitive abilities, and so reduced validity and maximized the level of adverse impact (against blacks) of the Written Tests. This is a major shortcoming because tests of non-cognitive abilities show less adverse impact than tests of cognitive abilities, while contributing to the validity of the selection process (as testified to by Defendant's experts, as I say above). In sum, important abilities, both non-cognitive and cognitive, were omitted from both the DCAS job analysis and the WTs.

⁶⁷Schemmer 290 line 4 - page 291 line 16, page 301 line 17 - page 302 line 5

⁶⁸Test Development Report for Exam 7029, page USA004814.

⁶⁹Test Development Report for Exam 7029, page USA004918.

⁷⁰The full report is numbered USA004811-USA004918.

⁷¹Bilingual abilities are also not to be found in the DCAS JAQ.

The importance of certain personal characteristics for successful performance as a Firefighter was made clear in a nationwide study conducted by the US Civil Service Commission (US CSC) in 1976 and published by the US CSC in 1977⁷² and then reported again in a 1988 in a textbook on job analysis. (In the 1970's the US CSC, now US OPM, was providing professional support to state and local merit systems in the area of personnel assessment and personnel selection, in support of merit system principles.) Based on a review of the literature and a job analysis study of over 100 fire departments nationwide, the US CSC report provided a test weighting plan for 20 abilities and characteristics. (The US CSC report also provided a list of existing tests that measure each of these 20 abilities/characteristics.) This report shows that the 5 highest-weighted (highest importance) abilities/characteristics are non-cognitive, personal or interpersonal in nature (i.e., responsibility, desire to learn, teamwork, activity (energy), getting along with people, see Table 2). None of these most highly important non-cognitive abilities are included in the Exams 7029 or 2043. Of the 20 abilities or characteristics that the US CSC included in its recommended test weighting plan, only 4 are cognitive: problem-solving ability, mechanical ability, verbal skills (both written and verbal), and math skills. Of these, DCAS measured only 1½ (problem solving and written communication). Thus Exams 7029 and 2043 covered only 7.5% of the abilities and characteristics that the US CSC report recommended for an entry-level Firefighter written test.

When asked in deposition what makes for a good Firefighter, Chief Nigro (the highest ranking uniformed officer in the FDNY) mentioned only non-cognitive abilities and characteristics, saying,

Page 71

- 6 A. [Nigro] What makes a good Firefighter?
7 That's – that's a good question. What I think
8 you know, I think someone who's reliable, honest,
9 physically fit, and someone who has a big heart,
8 both in being able to go the extra mile, and also
10 someone who has – cares about other people.

⁷²Personnel Research and Development Center of the US Civil Service Commission, Technical Study 77-B, dated December 3, 1976.

Table 2. US CSC Weighting Plan for Firefighter Exam⁷³	
Weight	Abilities and Characteristics
8.0	Responsibility
7.9	Desire to Learn
7.9	Teamwork
7.4	Activity (energy)
7.1	Getting Along with People
6.1*	Problem-solving Ability
6.0	Honesty
5.4	Visual Acuity
5.1	Mechanical Ability
5.0	Resistance to Stress
4.5	Quickness (physical)
4.3	General Body Coordination
4.2	Cleanliness
4.1	Dexterity (of limbs and fingers)
4.0**	Verbal Skills (written and verbal)
3.8	Construction Trade Interest
3.5	Physical Strength
3.2	Courage
1.5	Math Skills
1.1	Medical interests
100	Total

* DCAS measured this ability in Exams 7029 and 2043 (in bold above)

** DCAS measured the written part of this ability in Exams 7029 and 2043 (half in bold above)

⁷³Derived from Table 20 (page 94) of Bownas & Heckman (1976) *Job Analysis of the Entry Level Firefighter Position*. Washington, DC: Personnel Research and Development Center of the US Civil Service Commission.

The current federal job description system known as the O*Net identifies the following work-style areas⁷⁴ as being at the very important level for the job of Firefighter.⁷⁵ achievement/effort, adaptability/flexibility, analytical thinking, attention to detail, concern for others, cooperation, dependability, initiative, integrity, persistence, self control, social orientation, and stress tolerance.⁷⁶ DCAS did not test for these work-style areas in Exams 7029 and 2043 although DCAS expert, Dr. Schemmer, said it was possible to do so.⁷⁷ (See Attachment A for a reproduction of the relevant section of the O*Net web page for Municipal Firefighter.)

Job analysis studies done in various jurisdictions nationwide prior to the time of preparation of Exam 7029 have shown the importance of non-cognitive abilities for successful job performance as a Firefighter. For example, a multi-jurisdictional study done in 1979 revealed the importance of non-cognitive abilities such as: integrity, teamwork, emotional self-control, and dependability.⁷⁸ DCAS did not test for these non-cognitive areas in Exams 7029 and 2043.

Criterion-related validation studies done in various jurisdictions nationwide prior to the time of preparation of Exam 7029 have also shown the importance of non-cognitive abilities for successful job performance as a Firefighter. For example, a multi-jurisdictional study done in 1990 showed that "the interpersonal performance dimensions carried twice the weight of the technical dimensions in the raters judgements of overall [job] performance."⁷⁹ The simple correlations between test scores and overall job performance scores were .23 for memory and reading and .20 for responsibility and emotional stability.⁸⁰ That these numbers are so close means that

⁷⁴Personal characteristics that can affect how well a person performs the job of Firefighter.

⁷⁵Detailed summary for Municipal Firefighter downloaded (on 5/29/08) from the O*NET website: <http://online.onetcenter.org/link/details/33-2011.01#WorkStyles>

⁷⁶Among Firefighters, stress tolerance is related to abuse of alcohol and illegal drugs: Bacharach, Bamberger & Doveh (2008)

⁷⁷Schemmer page 300 lines 4-14.

⁷⁸"The Development and Validation of a Multijurisdictional Fire Service Entrance Examination for the State of Utah" (Table 9, page 12).

⁷⁹Sommerfeld, D. (1990) Firefighter Test Development and Criterion-Related Validation Project. Ann Arbor Michigan: Michigan Municipal League. (Page 6-6)

⁸⁰Sommerfeld, D. (1990) Page 6-18

cognitive and non-cognitive abilities were comparable in validity. DCAS did not test for responsibility and stability in Exams 7029 and 2043.

The importance of non-cognitive abilities for NYC Firefighters is further supported by DCAS Written Test 6019 (first administered in 2007) which measured the following 7 personal characteristics: adaptability, tenacity, integrity, work standards, resilience, coordination (with others), and establishing and maintaining interpersonal relationships.⁸¹ DCAS did not test for these personal characteristics in Exam 7029 and 2043.

Both job analysis and criterion-related validity studies done nationwide prior to the time of preparation of Exam 7029 support the validity of mechanical ability. For example, a multi-jurisdictional study done by the US OPM in 1980⁸² found that mechanical ability had the highest correlation with the overall criterion (.20), while reading ability correlated .02 and arithmetic ability correlated .12 with the overall criterion. DCAS did not test for mechanical ability in Exams 7029 and 2043. It is also noteworthy that DCAS did not test mechanical ability, a cognitive ability that could have been tested easily and which has been reported to be a valid cognitive ability for predicting the job performance of Firefighters,⁸³ and for which tests have been available for many years.⁸⁴

Even after the initial decision to use only Fleishman's cognitive ability areas, DCAS might have noticed some of these omissions if the job analysis were done a little differently. In the DCAS job analysis for Exam 7029, the ability list was not vetted by the Firefighters interviewed using the Field Interview Guide,⁸⁵ and there was no place on this form for incumbents to add additional characteristics or abilities that they thought were necessary.⁸⁶ Either of these might have revealed to the examiner that

⁸¹Notice of Examination for Firefighter Exam. No. 6019, page 2, paragraph 2.

⁸²Bullock, C. (1980) A Study of the Validity of Selected Tests for the Selection of Entry-Level Firefighters in Three Maryland Counties. Page 51.

⁸³The New York State Department of Civil Service examination for Firefighter includes mechanical reasoning questions. (Guide to Taking the Written Test for Firefighter Series (2002) New York State Department of Civil Service; Albany, New York, page 5.)

⁸⁴Schemmer page 263 lines 16-22

⁸⁵The Field Interview Guide did not contain any questions about the abilities that would be considered for inclusion in the examination, even though they were known at the time of the field interviews/observations. Pages USA004826 - USA004846.

⁸⁶Test Development Report for Exam 7029, page USA004864.

additional abilities should have been considered.

Related to this restricted range of abilities considered, the Test Development Report for Exam 7029 has two contradictory statements on the page where the preparation for the job analysis is discussed:⁸⁷

Conduct a new, comprehensive job analysis.

Use a Fleishman-based ability test. Among these abilities are Oral Comprehension, Write [sic] Comprehension, Oral Expression, and Written Expression.

The stated goal, a "comprehensive" job analysis is vitiated by the unsubstantiated decision to limit the test to cognitive Fleishman abilities.

2. Fleishman Abilities Not Understood

The list of Fleishman abilities (e.g., Spatial Orientation and Visualization) used in the job analysis for Exam 7029 are not easily understood by the lay person. The multiple problems this spawned contributed to the collection of job analysis data with unknown meaning.

a. Linking Panel Members Now Disagree with Their Own Ability Ratings

In deposition, _____, one of the 12 Linking Panel members who rated the importance of the Fleishman abilities, said that Written Expression was not important to the job of Firefighter.⁸⁸ This supports the current evaluation of _____ another Linking Panel member, that he rated Written Expression too high when he was a member of the Linking Panel.

Mr. _____ said his rating of importance of Visualization for the salvage task cluster was too high, in retrospect.⁸⁹ This Linking Panel member rated Written Expression as "Somewhat important" for Initial Response to Incidents/Driving, but in his deposition he said "a driver is not going to do much driving and writing at the same time",⁹⁰ and he would rate Written Expression as zero if he were to do the ratings again.⁹¹ Similarly, now he would rate Written Expression as "Not relevant" for the Ventilation task cluster, but as a Linking Panel member he rated Written Expression as "Somewhat

⁸⁷Test Development Report for Exam 7029, page USA004814.

⁸⁸"On a firefighter's job, it's not that important." (_____, page 43, lines 9-12).

⁸⁹_____ page 101, lines 19-25, page 102 lines 2-5.

⁹⁰_____ page 120, lines 8-10.

⁹¹_____ page 120 lines 21-24.

Table 3. Illustration of Multiplying versus Averaging Ratings				
Ability	JAQ Average	Linking Panel Average	Multiply	Average
Ability X	3	1	3	2
Ability Y	4	3	12	3.5
Difference	1	2	9	1.5

7. JAQ Ability Ratings Received Less Weight than Linking Panel Ability Ratings

The JAQ average ability importance ratings had little variability, with a range from the lowest to the highest average rating of less than .5 of a scale point for the average ratings of the 9 abilities tested on the 7029/2043 Written Tests. On the other hand, the Linking Panel average ability importance ratings had more variability, with a range of 1.1 scale points for the average ratings of the 9 abilities tested. This means the JAQ ability importance ratings (based on 192 people) had much less weight than the Linking Panel importance ratings (based on 12 people). This inequity in weighting the ratings of the JAQ and the Linking Panel was apparently unintended. It should have been noted and addressed in the Test Development Report, but it was not.¹⁴⁴ This is another example of the job analysis being done in a formulaic, thoughtless fashion rather than as an application of applied research, which requires dealing with the data collected in a thoughtful fashion.

8. Another Anomaly in the Combination of Ability Importance Ratings

DCAS made another logical error when it combined the two sets of ability importance ratings, those collected from 192 Firefighters who completed the JAQ and those from the 12 Linking Panel members. Under the DCAS method, the two importance ratings for each ability area were multiplied to get an overall importance rating for each ability.¹⁴⁵ The error arose due to the way DCAS assigned numbers to the 4 possible importance ratings. A four point importance rating scale was used for the JAQ and the Linking Panel, but different numeric values were assigned to the scale points: 1-4 for the JAQ, 0-3 for the Linking Panel (see Table 4).

¹⁴⁴DCAS recognized a very similar problem in combining the scores of the WT and the PPT (and dealt with it by standardizing), but ignored this problem in combining the scores of the JAQ and the Linking Panel.

¹⁴⁵Within ability this was done for each Task Cluster, and then the products for all the Clusters were averaged. (Test Development Report for Exam 7029, page USA004910).

comporting with these widely accepted principles of test writing that the quality of the test questions and the whole test is seriously compromised.

c. Reading Level of Test Overall Is Too High for the Job

My analyses show that the reading level for Written Tests 7029 and 2043 was too high for the job of Firefighter. I calculated the reading grade level for these tests and for the individual questions.¹⁶² There are several widely used formulas used to calculate reading grade level. The most appropriate formula to calculate the reading grade level for multiple-choice tests is the "SMOG" formula. Unlike other approaches to measuring reading grade level, the SMOG formula was developed to predict 100% comprehension of written passages. (The widely used Flesch approach was designed to predict 75% understanding.¹⁶³) I report reading grade levels based on the SMOG formula, unless otherwise noted. The reading grade level that is most appropriate varies depending on whether the reading material must be understood with or without assistance. For example, in a training environment where discussions with trainers or other trainees are allowed, a higher reading grade level can be tolerated. When a person is reading in an isolated or unassisted fashion, as when taking a civil service test, a lower reading grade level is important so that low reading skill does not obscure the ability being tested.

When written material is used in the training of Firefighters, both in the Fire Academy and in the fire station, the firefighters are allowed and even encouraged to ask questions of supervisors, trainers, or other colleagues.¹⁶⁴ For example, one Linking Panel member testified as follows:

Page 115

23 Q. If there is anything in the
24 written materials that someone in the
25 academy didn't understand, is there

Page 116

1 M.J. CARLIN
2 always an opportunity to ask questions
3 about the materials?

¹⁶²Readability formulas use various objective measures to predict the level of reading skill needed to understand the prose material at hand to a specified criterion level of understanding, and most describe that level of skill in terms of the "reading grade level" needed to comprehend the material to that criterion level of understanding. Most such formulas rely on two measures, one for sentence and one for word difficulty.

¹⁶³As described, for example, in DuBay (2007) page 109.

¹⁶⁴For example, Barretta, page 123 line 21 - page 124 line 21, and page 125 line 8 - page 126 line 6; and Kearns page 124 line 11 - page 125 line 18; Barron page 116 line 4 - line 14.

- 4 A. Absolutely.
5 Q. And you can get an oral
6 explanation of what's in the materials?
7 A. Yes.
8 Q. As far as you remember, are
9 there also visual demonstrations of
10 what's taught in the written materials?
11 A. Yes.

This is a stark contrast to the conditions for taking a written multiple-choice test where talking (and asking questions) is forbidden. Therefore, the reading level requirements of written test questions should be below the reading level of material Firefighters may be given to read at the Fire Academy or in the fire station where assistance is readily available.

My reading grade level analysis shows that the average reading grade level for 7029 and 2043 questions were 12.3 and 12.8.¹⁶⁹ Both the 7029 and 2043 Written Tests contained many questions with high reading grade level, well into college and even graduate school levels. Since the DCAS Minimum Requirements for Firefighter allow people to take the test who have only a High School education, I tallied the number of questions using 12th grade reading level as a cutoff: 50 of the 85 questions on the 7029 Written Test and 56 of the 86 questions on the 2043 Written Test had reading grade levels above the 12th grade. (See Table 7 for a summary of question reading grade levels.)

For the sake of comparison, I provide similar readability information for two other written tests for Firefighter. One Firefighter test used widely is the Merit Employment Assessment Services (MEAS) test authored by Dr. Thomas Tyler. The MEAS readability analysis (see Attachment C) shows that his test has a 7th grade reading grade level (using the Flesch formula), a figure which I adjust up by 3 reading grades to be comparable to the SMOG readability analysis. This yields a SMOG-comparable reading grade level of 10th grade, more than 2 years lower than the comparable reading grade levels for Written Tests 7029 and 2043. Clearly the reading level of WT 7029 and 2043 are higher than needed for a Firefighter exam.

¹⁶⁹That this is not driven by the questions measuring Written Comprehension is indicated by the average reading grade levels of 12.5 and 12.7 for the questions for abilities other than Written Comprehension or Memorization. Parenthetically I note that the DCAS readability analysis for 7029 contains a handwritten annotation that says, "Readability level without paragraphs and headings." (DCAS 004492) Given that the test instructions tell the test taker to answer the questions based on the accompanying paragraphs, omitting the paragraphs from the readability analysis yields an artificially low estimate of the reading level demands of the WTs.

Table 7. Summary of Question Reading Grade Levels for Four Tests					
	Whole Exam* Reading Grade Level	Lowest Question Reading Grade Level**	Highest Question Reading Grade Level	Mean Question Reading Grade Level	Number of Questions with Reading Grade Level above 12th Grade
7029 Exam	12.9	3.1	22.2	12.3	50
2043 Exam	12.7	3.1	18.9	12.8	56
Tyler Exam***	10.6	n.a.	n.a.	n.a.	n.a.

* Reading grade level was computed using the correct answers only. Incorrect answers were not considered.

** The questions with the lowest reading grade levels all tested Memorization

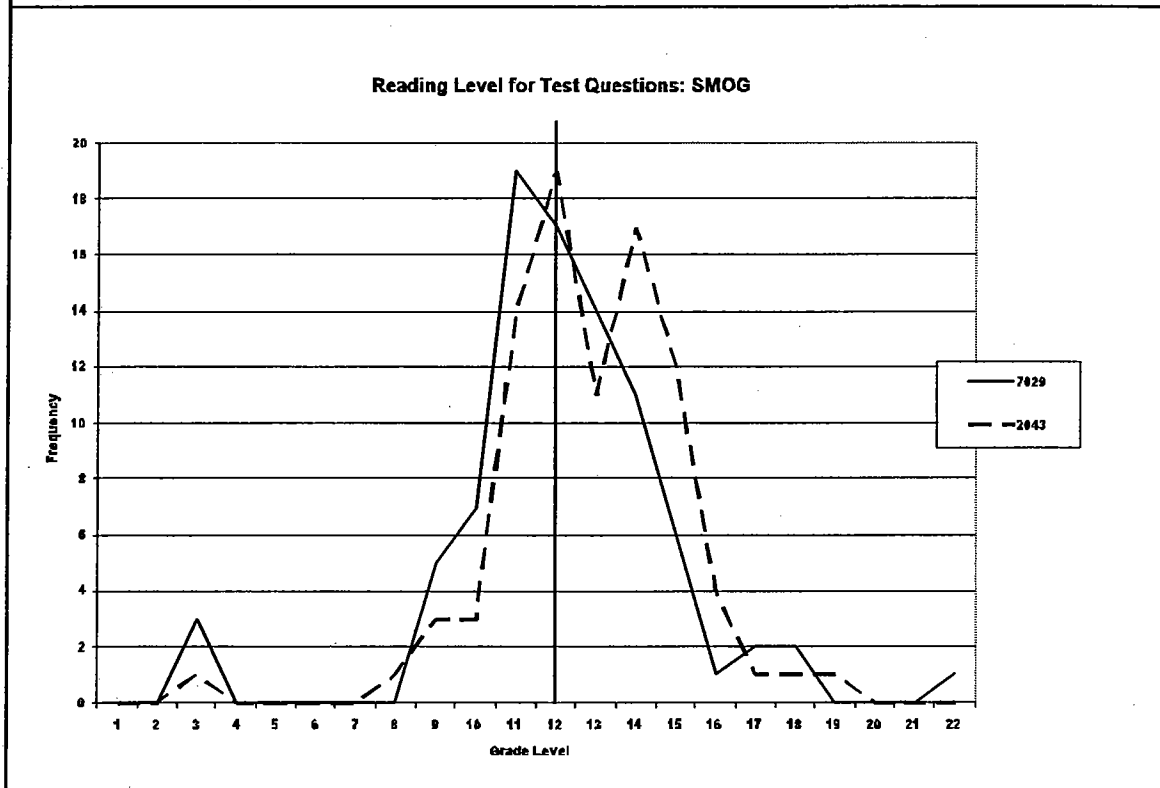
*** This value is estimated from the Flesch analysis done by Tyler by adding 3 reading grade levels to his reading grade level.¹⁶⁶

d. Reading Level of Many Individual Questions Too High for the Job

Often a readability analysis of test questions is done on the whole test rather than on individual test questions, since individual test questions are usually very short, and short reading passages can yield spuriously high or low reading levels. However, over half of the questions on both Written Tests 7029 and 2043 contain more than 100 words, so that concern is less relevant for these questions.¹⁶⁷ A majority of the questions on both the 7029 and 2043 Written Tests were over the 12th grade reading level (see Table 7, Figure 1).

¹⁶⁶Often the SMOG reading grade level is from 2 to 3 years higher than the Flesch.

¹⁶⁷Written Test 7029 contains 11,844 words, and Written Test 2043 contains 11,517 words (omitting headings and tables, but including instructions and wrong answers, with number of spaces taken as a count of the number of words).

Figure 1. Reading Grade Level for Questions in WT 7029 and 2043

In summary, the DCAS 7029 and 2043 Written Tests for Firefighter contained a majority of questions written at a reading grade level too high for the job. Many applicants did not have a fair opportunity to show their abilities, for lack of a reasonable opportunity to understand many of the questions on the 7029 and 2043 Written Tests.

e. High Reading Level Violated the APA Standard on Fairness

The *APA Standards* has a section entitled “Fairness in Testing and Test Use.” One of the “Standards” in that section reads as follows:

“In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct.

Comment: When the intent is to assess ability in mathematics or mechanical comprehension, for example, the test should not contain unusual words or complicated syntactic conventions unrelated to the mathematical or

1. PPT Weight is Zero for Most Test Passers

The intended weight for the PPT was 50% for both Exams 7029 and 2043.²²⁵ However, for most of the test takers the PPT had no effect on their ranks at all. Over 70% of the test takers for each exam scored 100% on the PPT (see Table 9). For these test takers the final rankings were determined solely by the Written Test score and special point credits.

Table 9. PPT Passing Score Distribution for Exams 7029 and 2043		
Score	7029 Number/ Percent	2043 Number/ Percent
75	549 9.7%	535 8.2%
87.5	1,139 20.1%	1,048 16.1%
100	3,974 70.2%	4,943 75.7%

2. Equal Weighting of WT and PPT is Unsupported

The decision to use a WT and PPT, weighted equally, was made prior to collection of any job analysis data.²²⁶ This timing, together with the decision not to consider physical abilities in the job analysis,²²⁷ means that the job analysis could provide no support for the validity of the choice of a WT and a PPT nor their relative weight. Thus there is no validation support for the equal weighting of the WT and PPT.

In addition, Mr. Alexander testified that DCAS has data collected in 1992 and again in 2005 that show that the weight for the PPT and WT should be 55% and 45%, respectively.²²⁸

3. PPT z-Score Weighting Disadvantages Blacks

The DCAS method worked to the disadvantage of blacks for a subtle reason related to

²²⁵As stated on the DCAS Notice for Examination for 7029 and 2043 (pages USA002575 and USA002572).

²²⁶The Test Development Report for Exam 7029, page USA004814.

²²⁷The Test Development Report for Exam 7029, page USA004814.

²²⁸Alexander page 98 line 8 - page 99 line 4.

truncation of scores. The reason for this is difficult to explain concisely, but this problem in scoring impacts every black test passer who did not get a perfect score on the PPT. The problem is that DCAS did not collect scores below 62.5 on the PPT. That throws off the z score weighting, as follows. The z score weighting uses and depends on both the mean and the standard deviation of the PPT raw score. When DCAS calculated the mean PPT using only the scores they recorded (i.e., 62.5, 75, 87.5, and 100) the mean PPT raw score is about 87.5.²²⁹ However, some of the people whose PPT scores were recorded by DCAS as 62.5 were not allowed to complete the PPT test, and some of them would have scored lower than 62.5. If the mean were to be calculated on the full (real) PPT score distribution (including people with scores of 0, 12.5, 25, 27.5, and 50) the mean would be less than 87.5. That would mean that people with a raw PPT score of 87.5 would now be above the (new lower) mean. That would improve their z-score. If the standard deviation were calculated on the full PPT distribution it would be greater, since it would be based on a wider range of possible PPT scores (0 to 100 as compared to 62.5 to 100). The effect of both of these on the z- scores for passers would be to increase the z-scores for those people with a PPT raw score of 87.5 (since 87.5 would be above the mean) and decrease the positive z-scores for those people with a PPT raw score of 100. This would work to the advantage of blacks over whites on Exam 7029 since proportionally more blacks scored 87.5 and proportionally fewer blacks scored 100, as compared with whites. The z-scores for those with a PPT raw score of 75 would be either positive or less negative. This would help blacks since proportionally more blacks than whites achieved a score of 75. In short, the flawed z-score methodology worked to increase adverse impact on blacks on Exam 7029. The same would be seen with Exam 2043 where the mean PPT raw score was calculated by DCAS to be about 89.5.²³⁰ As for Exam 7029, the scores for blacks on Exam 2043 would benefit more, relative to whites, from calculating the z-scores based on the full PPT distribution.

4. Passing Points for the Written Tests Were Not Validated

The merit system in this country was established to combat patronage and assure that people hired are competent to perform the job. Often, public sector employers set the passing score at a point that attempts to assure competence. There is no indication that DCAS did that. The passing points for the WT for Exams 7029 and 2043 are quite different (84.705 and 70.0 respectively), even though the two Exams are roughly comparable in difficulty. The average scores on the two Written Tests differ by about 2 points whereas the passing points for the two tests differ by about 14 points (as shown in Table 10).

²²⁹The DCAS Explanation of Test Scores for Exam 7029 reports the PPT mean as 87.262 (page USA004935).

²³⁰The DCAS Explanation of Test Scores for Exam 2043 reports the PPT mean as 89.494 (page USA0004931).

takers who left one or more of these 10 questions blank was analyzed for whites and blacks.²⁰² On the 7029 exam, 0.2% of whites left one or more of these ten questions blank; 1.1% of blacks left one or more of these ten questions blank. On the 2043 exam, 0.4% of whites left one or more of these ten questions blank; 2.7% of blacks left one or more of these questions blank. This difference between whites and blacks for blanks in questions 16-25 was tested for statistical significance. On both exams, blacks had statistically significantly more blanks in questions 16-25 than whites.²⁰³ One plausible explanation of this difference is that it reflects a difference in test taking sophistication, sometimes referred to as less testwise, since it is to the benefit of the test taker to answer all the questions on a test such as WT 7029 or 2043. It may also reflect unnecessarily difficult wording or jargon in the questions.

d. Blacks Left More Questions Blank at the End of the Written Tests

For each exam, the number of test takers who left the last ten questions blank was analyzed for whites and blacks. On both exams, blacks had significantly more blanks in the last 10 questions than whites. On the 7029 exam, less than 1% of whites left one or more of the last ten questions blank while 3% of blacks left one or more of the last ten questions blank. On the 2043 exam, less than 1% of whites left one or more of the last ten questions blank while 3% of blacks left one or more of the last ten questions blank. This difference between whites and blacks for blanks in the last 10 questions was tested for statistical significance. On both exams, blacks had statistically significantly more blanks in the last 10 questions than whites.²⁰⁴ It may be that the high reading level and similar flaws in test development made it harder for blacks, on average, to finish the test. Since the B-W difference in mean score on the WT is about 8 points for both 7029 and 2043, any aspect of the WT that might have contributed even a one point difference between white and black test takers is of great practical import.

e. These Tests Were Speeded But Were Not Intended to Be

With respect to time limits, multiple-choice employment tests are either power tests, with a generous time limit that allows all test takers to answer all the questions, or speed tests, with a short time limit due to the nature of the test or the testing situation, in which case many or all test takers are expected not to finish. The WT for 7029 and 2043 were intended to be power tests. I say this for several reasons. First, that is the most typical approach to testing the abilities on these tests. Second, the time limit was 3.25 hours, which is a typical time limit for a power test. Third, the DCAS Instructions for Monitors for WT 7029 say that fingerprinting of test takers will take place during

²⁰²Questions 16-25 were chosen because they appear near the start of the test, but after the Memorization questions.

²⁰³7029: $F = 23.1$, $p < 0.001$; 2043: $F = 50.2$, $p < 0.001$

²⁰⁴7029: $F = 84.8$, $p < 0.001$; 2043: $F = 65.6$, $p < 0.001$

EXHIBIT AA

1
2 UNITED STATES DISTRICT COURT
3 EASTERN DISTRICT OF NEW YORK
4 Civil Action No. 07-CV-2067

-----X
5 THE UNITED STATES OF AMERICA,
6 Plaintiff,

7 and

8 VULCAN SOCIETY, INC., for itself and on
9 behalf of its members, CANDIDO NUNEZ,
10 MARCUS HAYWOOD and on behalf of a Class
11 of All Others Similarly Situated,

12 Plaintiff-Intervenors,

13 - against -

14 CITY OF NEW YORK, FIRE DEPARTMENT OF THE
15 CITY OF NEW YORK, NEW YORK CITY
16 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
17 SERVICES, MAYOR MICHAEL BLOOMBERG and
18 NEW YORK CITY FIRE COMMISSIONER NICHOLAS
19 SCOPETTA, in their Individual and
20 Official capacities,

21 Defendants.
-----X

22 271 Cadman Plaza East
23 Brooklyn, New York
24 July 8, 2008
25 9:55 a.m.

26 CONTINUED DEPOSITION of
27 Non-Party Witness, FREDERIC MARK SCHEMMER,
28 pursuant to Federal Rules of Civil
29 Procedure, held at the above place, date
30 and time, before Alice Schulman, a Notary
31 Public of the State of New York.

1 M. Schemmer

2 Q. Do you agree that even if you
3 have a valid test, the test still may be
4 used in such a way that the validity is
5 weakened or destroyed?

6 A. I would phrase that slightly
7 differently.

8 Q. How would you phrase it?

9 A. I would say that because
10 validity is as much a function of the
11 process and application of the test as is
12 the test itself, that certainly validity
13 can be affected by an improper application
14 or inference as validity doesn't follow
15 the test around.

16 Q. Do you agree with this
17 statement, no matter how valid the exam,
18 it is the cutoff score that ultimately
19 determines whether a person passes or
20 fails. A cutoff score unrelated to job
21 performance may well lead to the rejection
22 of applicants who are fully capable of
23 performing the job.

24 A. As stated, I would agree with
25 that.

1 M. Schemmer

2 Q. Let me ask this, you recognize
3 one of the things that you agreed with as
4 coming from the APA standards, correct?

5 A. I believe so.

6 Q. When the APA standards refer to
7 interpretations of test scores required by
8 proposed uses, what does that mean?

9 A. Essentially, I would take that
10 to refer to the particular application of
11 the test scores in whatever personnel
12 selection process in this instance.

13 Q. Let me ask this, what
14 interpretation of test scores is required
15 by use of a test as a pass/fail screening
16 device with a cutoff score of 84.705?

17 A. I apologize, could you repeat
18 the question?

19 Q. Sure. My question is what
20 interpretation of test scores is required
21 by use of a test as a pass/fail screening
22 device with a cutoff score of 84.705?

23 A. The fundamental interpretation
24 would be that individuals scoring higher
25 than that cutoff had a greater likelihood

1 M. Schemmer

2 of successfully performing the job or
3 performing the job at a higher level of
4 competency than those lower than that test
5 score.

6 Q. Let me ask this, what
7 interpretation is required by use of a
8 test for rank order selection?

9 A. Essentially the same base
10 interpretation with the more explicit
11 expectation of a monotonic relationship
12 between test scorer or test performance on
13 the one hand and outcome measure on the
14 other hand.

15 Q. Just because test scores on a
16 particular test have a correlation with
17 job performance, that doesn't necessarily
18 indicate that there's a monotonic
19 relationship; is that correct?

20 A. In general, it would indicate a
21 monotonic relationship. Regression or
22 correlation coefficients most typically
23 reflect a linear relationship which is
24 monotonic.

25 Q. But not necessarily, correct?

1 M. Schemmer

2 A. A linear relationship is
3 monotonic. A monotonic relationship is
4 not necessarily linear.

5 Q. Okay. Do you agree with this
6 statement, where composite scores are
7 developed, the basis and rationale for
8 arriving at the composites should be
9 given?

10 A. I would in principle concur.

11 Q. Do you agree that in such a
12 situation the rationale and supporting
13 evidence must pertain directly to the
14 specific score or score combination to be
15 interpreted or used?

16 A. One more time.

17 Q. Do you agree that where
18 composite scores are used, the rationale
19 and supporting evidence must pertain
20 directly to the specific score or score
21 combination to be used?

22 A. I would agree with that.

23 Q. In your report on page 29, in
24 fact, the last sentence on page 29 you say
25 changes to specific content elements

1 M. Schemmer

2 characteristics, did it?

3 A. I don't recall that it did.

4 Q. Then why would you say that you
5 don't agree that the job analysis the City
6 did for exam 7029 overemphasized cognitive
7 skills?

8 A. Perhaps to make the, parsing to
9 make the statement that it overemphasized
10 cognitive skills, I would want some sort
11 of definitive information suggesting that,
12 I'll say absent component whatever they
13 are of the sort on page 15 of this
14 document or others, that information, had
15 they been included in that job analysis,
16 such other things would have shown a
17 similar level of importance or frequency
18 or depending on whatever sort of ratings
19 were used.

20 Q. If I told you that the job
21 analysis that was done for exam 6019 found
22 that physical abilities and personal
23 attributes were as important as cognitive
24 skills for the New York firefighter job,
25 and also found that the job had not

1 M. Schemmer

2 changed from the early 90's, would you
3 then agree the job analysis the City
4 conducted for exam 7029 overemphasized
5 cognitive skills?

6 A. I would certainly come a great
7 deal closer to agreeing with that, yes.

8 Q. Do you agree that increasing the
9 job analysis domain to capture the full
10 range of KSAs will likely decrease group
11 differences, reduce group differences?

12 A. At a general level, I do not
13 believe that is necessary, would
14 necessarily be the outcome.

15 Q. I asked if it was likely.

16 A. I do not believe it would be
17 likely.

18 Q. Okay. If you look at page 20 of
19 Exhibit 663.

20 A. Page 20?

21 Q. Page 20. It has a heading Job
22 Analysis Support for Measuring Multiple
23 Dimensions. Now, looking at this chart on
24 page 20, in the job analysis that they're
25 reporting here, which category of ability

1 M. Schemmer

2 with practical skills.

3 Q. You testified earlier about the
4 Barrett study, do you recall that
5 testimony?

6 A. Yes, sir.

7 Q. And you agreed that Barrett had
8 tested for mechanical, or Barrett's study
9 referred to the use of mechanical skills
10 as one of the criteria that was used in
11 his meta-analysis; is that right?

12 A. The Barrett study categorized or
13 broke the tests into two categories that
14 they described or labeled as mechanical
15 and cognitive.

16 Q. At the time that 7029 and 2043
17 were created and administered, the City
18 could have tested for mechanical skills on
19 those exams?

20 A. Mechanical skills, aptitudes and
21 the like, certainly tests of those have
22 been around for quite awhile.

23 Q. In your testimony last week, you
24 mentioned that you had received and looked
25 over a document that was referred to as a

1 M. Schemmer

2 Q. Cleanliness, important to
3 firefighters?

4 A. Yes.

5 Q. Medical interest, which I think
6 you used as a criteria, important to
7 firefighters?

8 A. Yes.

9 Q. According to this document,
10 there were validating tests back in the
11 70's for all of these KSAs. Could these
12 all have been tested for in 1999 and 2002
13 when 7029 and 2043 were administered?

14 A. Certainly tests of those
15 constructs or categories were available
16 during that time frame.

17 Q. Are you familiar with a company
18 called Previsor?

19 A. Generally, yes.

20 Q. Is that where Dr. Bobko has a
21 relationship?

22 A. I believe he either does or did
23 have an advisory relationship with them.
24 I don't know if that's current.

25 Q. I think he testified that he was

1 M. Schemmer

2 on their technical advisory committee,
3 does that sound right to you?

4 A. Yes.

5 MR. LEVY: Could we mark this
6 66-A, please.

7 [The document was hereby marked
8 as Plaintiff-Intervenors' Exhibit 66-A
9 for identification, as of this
10 date.].

11 Q. And I'm handing you a document
12 previously marked 66. Do you know what,
13 do you know of Previsor and the work they
14 do?

15 A. At a general level, yes, sir,
16 they are in a remote sense a competitor in
17 the assessment arena to some of
18 ChoicePoint's assessment area.

19 Q. You wouldn't bad mouth their
20 assessment here?

21 A. Not at all.

22 Q. The item that's been marked as
23 Plaintiff's Exhibit 66 says that,
24 advertises that Previsor does personality
25 assessments, evaluating factors such as

1 M. Schemmer

2 emotional control, stress tolerance, self
3 confidence, adaptability, independence,
4 responsibility, initiative
5 competitiveness, social awareness, empathy
6 and influence.

7 Are those all items that can be
8 tested for on a written test?

9 A. They are certainly within
10 assessments or paper and pencil
11 assessments of those constructs.

12 Q. And Dr. Bobko testified that the
13 company maintains that these are valid
14 assessment measures. Do you have any
15 reason to believe they are not or would
16 not be?

17 A. No, I do not.

18 Q. Further down there is a category
19 called behavioral that involves biodata
20 instruments used by Previsor. Have you
21 used biodata instruments?

22 A. Some work that I have done in
23 the past involve biodata instruments, yes.

24 Q. And do you know if they were
25 valid assessment instruments?

1 M. Schemmer

2 A. Certainly in at least one study
3 that I can recall the biodata component
4 showed evidence of validity.

5 Q. The assessments referred to on
6 Plaintiffs' Exhibit 66 are all evaluation
7 measures of the type that could have been
8 utilized in a paper, pencil test in 1999
9 and 2002 when 7029 and 2043 were
10 administered?

11 A. I would think so.

12 Q. Let's take a look at the
13 document that's been marked 66-A, also
14 from Previsor. Did I give you this yet?

15 A. I think so.

16 Q. If you look at some of the jobs
17 and the KSAs that were measured, there's
18 not a job I should say for firefighter
19 here. It's not one of the jobs that they
20 list, but I think some of the jobs that
21 are listed here have some overlap, and I
22 will ask you about those.

23 Do you think on page 3 the job
24 that's called general skilled job that
25 would call material handlers, maintenance

1 M. Schemmer

2 orientation, dependability and
3 conscientiousness are items that are
4 tested for on paper and pencil tests?

5 A. Yes, sir.

6 Q. And could have been tested for
7 on the 7029 and 2043 exams?

8 A. Yes, sir, exactly the same
9 answer I gave to the same set of KSAs in
10 the general skill construction.

11 Q. Would you think that well
12 developed tests to assess these KSAs that
13 we've talked about that are referred to on
14 66-A could add validity to an exam for
15 firefighter?

16 A. I would say they certainly have
17 the potential to do so, yes, sir.

18 Q. In your opinion, do you think
19 that these measures would have a lower
20 black-white difference in results, a
21 different d, a lower d than a purely
22 cognitive test would?

23 MR. FRAENKEL: Objection as to
24 form, but you can answer.

25 A. Certainly the literature

1 M. Schemmer

2 suggests that these sorts of KSAs, I will
3 broadly label them personality
4 characteristics, typically show small
5 letter d's or lesser differences between
6 black and white job applicants or
7 incumbents, yes.

8 Q. Have you read Dr. Bobko's --

9 MR. LEVY: Withdrawn.

10 Q. I'm going to show you a document
11 that's previously been marked
12 Plaintiff-Intervenors' 65. These are the
13 O&NET KSAs for firefighter. Have you seen
14 this before?

15 A. I am familiar with O*NET. I
16 don't recall this specific page or list
17 offhand.

18 Q. Well, why don't you take a look
19 at it for a minute. I'm going to ask you,
20 perhaps without going through all of these
21 individually, whether these are all KSAs
22 that can be tested for on a written paper
23 and pencil test and could have been tested
24 for at the time that 7029 and 2043 were
25 administered.

1 M. Schemmer

2 A. (Witness complies.) Is there a
3 question outstanding?

4 Q. I just wanted you to look at it
5 with some advance advice that there would
6 be a question, and I think I told you what
7 the question is but I'll restate it.

8 It's the same question I was
9 asking you before about the Previsor
10 KSAs. Are these all KSAs that can be
11 tested for on a paper and pencil test?

12 A. I would suggest the majority
13 certainly have paper, pencil test
14 implementations.

15 Q. Are there particular ones that
16 don't, to your knowledge?

17 A. I'm hard pressed offhand on my
18 part to think of a paper, pencil test of
19 innovation, for example, towards the
20 bottom of the list.

21 Q. Okay. Any others?

22 A. There is certainly some debate
23 as to whether leadership is something that
24 can be systematically assessed with a
25 paper, pencil test.

1 M. Schemmer

2 Q. All right. Any others?

3 A. Self-control is one that, again,
4 depending on the particular
5 implementation, could be difficult in a
6 pre-employment situation since many
7 measures of self-control would potentially
8 step into difficulties with ADA.

9 Q. But could you test for
10 self-control, for example, in the way that
11 the Stanton test tests for dishonesty
12 which is, I guess, involves self-control?

13 A. I would buy that some
14 definitions of self-control or some
15 implementations might be testable
16 assessable with a written test.

17 Q. Would you say with respect to
18 these that you would expect that they
19 could, the use of them in an exam would
20 add to validity?

21 A. They certainly would have the
22 potential to do so, yes, sir.

23 Q. Would you also say that they,
24 most of these would be likely to have
25 lower adverse impact than purely cognitive

1 M. Schemmer

2 assessments?

3 A. Broadly or generally speaking, I
4 would anticipate they would have lesser
5 impact than cognitive assessments.

6 Q. Are these all apart from the
7 ones that you didn't think were testable,
8 KSAs for which there were assessments in
9 1999 and 2002 which could have been used
10 in the 7029 exam or the 2043 exam?

11 A. I suspect so, yes, sir.

12 MR. LEVY: It's a quarter to one
13 and I'm perfectly happy to keep going
14 but -- off the record.

15 (Discussion off the record.)

16 (Whereupon at 12:48 p.m. a
17 luncheon recess was taken.)

18 AFTERNOON SESSION.

19 (Time noted: 1:37 p.m.)

20 (Mr. Reese left the hearing room
21 and did not return.)

22 BY MR. LEVY:

23 Q. We are back on the record. Dr.
24 Schemmer, I've asked you to look at
25 exhibit, Plaintiffs' Exhibit 149 which is

1 M. Schemmer

2 Q. Which ones would not?

3 A. As one example, explosive
4 strength is, the physical abilities in the
5 Fleishman context, an example of one that
6 we very rarely find important to job
7 performance outside of the sports domain.

8 Q. Do you know why the reference on
9 the right says that Fleishman and
10 Quaintance, well O*NET, important
11 abilities for firefighters, they seem to
12 have rated it important?

13 (Mr. Ling entered the room.)

14 A. Point of clarification,
15 Fleishman and Quaintance would be the
16 reference source for the source of that
17 ability as opposed to the individuals who
18 wrote the O*NET documents themselves.
19 With many abilities, explosive strength is
20 a very good example of one. Without
21 careful explanation and delineation
22 between the abilities raters, typically
23 incumbents will often over focus on the
24 label, for example, explosive strength and
25 sort of generalize and not distinguish

309

1 M. Schemmer

2 between or amongst specific abilities.

3 That said, by all appearances,
4 whatever raters went into were fed the
5 O*NET documents for firefighters
6 apparently rated that as important.

7 Q. Doesn't it seem for, logically
8 to be important that a firefighter having
9 to burst into a burning building or grab a
10 child or a person in danger or break down
11 a door after climbing up two flights with
12 80 pounds, or whatever the weight is of
13 equipment, might be very much helped by a
14 good ability to exercise explosive
15 strength?

16 A. I wouldn't say it's irrelevant
17 by any stretch. What I'm suggesting from
18 my perspective and experience, and as a
19 particular example with firefighters, it
20 would be substantially less important than
21 static strength or dynamic strength.

22 Q. What else on this list would you
23 not rate as important? Referring to the
24 list again, it starts at 215 and goes
25 through 228.

EXHIBIT BB

1

2 UNITED STATES DISTRICT COURT
3 EASTERN DISTRICT OF NEW YORK
4 Civil Action No. 07-CV-2067

-----X

5 THE UNITED STATES OF AMERICA,
6 Plaintiff,

7 and
8 VULCAN SOCIETY, INC., for itself and on
9 behalf of its members, CANDIDO NUNEZ,
10 MARCUS HAYWOOD and on behalf of a Class
11 of All Others Similarly Situated,
12 Plaintiff-Intervenors

8

-against-

9

10 CITY OF NEW YORK, FIRE DEPARTMENT OF THE
11 CITY OF NEW YORK, NEW YORK CITY
12 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
13 SERVICES, MAYOR MICHAEL BLOOMBERG and
14 NEW YORK CITY FIRE COMMISSIONER NICHOLAS
15 SCOPETTA, in their Individual and
16 Official capacities,
17 Defendants.

13

-----X

14 February 21, 2008
15 10:01 a.m.

15

16 Deposition of CATHERINE S. CLINE, taken
17 by the Plaintiff, pursuant to Notice, at the
18 offices of the United States Attorney for the
19 Eastern District of New York, 271 Cadman Plaza
20 East, New York, New York, before David Levy,
21 CSR, a Notary Public of the State of New York.

22

23

24

25

1 Cline

2 looked at?

3 A. In re -- talking with -- in reviewing

4 documents prior to this deposition, I was made

5 aware of some documents by the attorneys that I

6 was not aware of previously.

7 Q. Now, so on your own, to prepare for

8 today's deposition, you looked, just looked back

9 at job analysis 6019, is that correct?

10 A. That's correct.

11 Q. And other than counsel, did you speak

12 with anyone to prepare for the deposition?

13 A. No.

14 Q. Now, you mentioned speaking with

15 counsel and being given documents to review. How

16 many times did you meet with counsel in

17 preparation for this deposition today?

18 A. Once.

19 Q. And when you're talking about

20 counsel, are you speaking only of Mr. Sample?

21 A. Mr. Sample and Mr. Frankel.

22 Q. Was anyone else present at the

23 meeting?

24 A. No.

25 Q. When did it occur?

1 Cline

2 A. Tuesday.

3 Q. Last Tuesday?

4 A. Two days ago. Today is Thursday,

5 right?

6 Q. Right.

7 A. Tuesday. The day before yesterday.

8 Q. How long was the meeting?

9 A. Two hours.

10 Q. And you said that you were given

11 documents that you had never seen before?

12 A. I was shown documents that I had

13 never -- never seen before. I wasn't given them.

14 Q. I see. Were you shown any documents

15 that you were familiar with?

16 A. One I was shown, which was the job

17 analysis report that was written by Matt

18 Morrongiello for one of the previous exams, I had

19 looked at before.

20 Q. Any other documents that you had seen

21 before?

22 A. No.

23 Q. What documents were you shown that

24 you were not familiar with?

25 A. I had never seen the document that

1 Cline

2 Q. Okay.

3 A. Of the test, or the subtest.

4 Q. Anything else?

5 A. I can't think of anything.

6 Q. Okay. So in terms of preparing the
7 test development report, my understanding from
8 your testimony is that you are the only person
9 preparing the report and Dr. Bobko has offered
10 you some comments and edits, is that correct?

11 A. That's correct.

12 Q. Is there anybody else involved?

13 A. No.

14 Q. Now, what is the highest level of
15 education you've attained?

16 A. I have a Ph.D.

17 Q. And from where did you receive this?

18 A. Columbia University.

19 Q. What was your major?

20 A. Psychology, measurement and
21 evaluation.

22 Q. When did you get your Ph.D.?

23 A. 1979.

24 Q. Okay. I assume you also have a
25 Master's degree?

1 Cline

2 through Columbia as part of your larger Ph.D.?

3 A. Yes.

4 Q. You also have a BA, I see?

5 A. Yes.

6 Q. And where did you receive your BA
7 from?

8 A. From Barnard College, Columbia
9 University.

10 Q. In what field?

11 A. Psychology.

12 Q. All right. And when did you get your
13 BA?

14 A. 1972.

15 Q. Do you have any other post-secondary
16 degrees?

17 A. No.

18 Q. Do you have any professional
19 certifications?

20 A. I have a license -- I'm licensed as a
21 psychologist in New York State.

22 Q. Anything else?

23 A. I've been accepted as an expert
24 witness in the areas of job analysis, test
25 development, adverse impact analysis, by a number

1 Cline

2 of different courts.

3 Q. You say job analysis, test

4 development, what was the other?

5 A. Adverse impact analysis.

6 Q. Anything else? Were you accepted as

7 an expert in any other fields?

8 A. No.

9 Q. Have you published any books,

10 articles or papers in the area of test

11 development or validation?

12 A. I did some papers that were on my

13 resume when I was first out of graduate school,

14 but I haven't really published anything since.

15 Q. How about in the area of job

16 analysis?

17 A. Not that I can remember. I don't

18 have anything major published in anything.

19 Q. Statistics?

20 A. No.

21 Q. How many job analyses have you

22 conducted?

23 A. Maybe a hundred.

24 Q. Okay. How many of those job analyses

25 are for public safety positions, would you

1 Cline

2 A. It wasn't a position. It was a -- it
3 was sort of like a self-informed -- to help
4 people -- more sort of like a personality
5 inventory. It was to try and help them know more
6 about themselves.

7 Q. Just part of the course you were
8 teaching?

9 A. No. It was just for -- it was just
10 for someone who was a friend-of-a-friend type
11 thing.

12 Q. Okay. Now, are you a member of the
13 American Psychological Association?

14 A. Yes, I am.

15 Q. How long have you been a member?

16 A. Probably since 1979 or '80.

17 Q. Do you hold any positions in the APA,
18 other than member?

19 A. No.

20 Q. Have you ever?

21 A. No.

22 Q. Are you familiar with a publication
23 called the Standards For Education and
24 Psychological Testing that was published by the
25 APA in 1999?

1 Cline

2 A. Yes.

3 Q. And have you read them?

4 A. Yes.

5 Q. And how do you use them as a
6 reference in developing or validating a test?

7 A. Again, pretty much what I said for
8 the standards. It's not like there's anything in
9 there that I wasn't acquainted with before they
10 wrote them. It's more to justify why I did
11 something or why something that was done is in
12 conformance with the standards, or is not
13 arbitrary and capricious.

14 Q. Are you a member of any other
15 professional organizations related to test
16 development or validation?

17 A. I'm a member of -- well, I'm maybe
18 not a member right now, the Metropolitan
19 Association of Applied Psychology.

20 Q. How long have you been a member of
21 this organization?

22 A. Twenty-five years.

23 Q. Have you held any leadership
24 positions in this organization?

25 A. I have been president,

1 Cline

2 similar to the physical test we had already.

3 Q. What do you mean by the PPT is or was
4 a competitive test?

5 A. The physical performance test?

6 Q. Correct.

7 A. They rank-ordered people in terms of
8 their performance on it.

9 Q. Okay. Can I ask, what was meant by
10 "feasibility" in terms of when it was said in
11 your report that you were trying to explore the
12 feasibility of using the CPAT?

13 A. I wanted to go talk to the Police
14 Academy and in fact, I did go and talk to the
15 police academy and I talked to the person who was
16 in charge of the police academy and who was in
17 charge of physical training for the recruits.

18 And one of the things I found out was that they
19 were using the CPAT as an exit-level measure of
20 physical fitness in the academy.

21 And I also found out that they were
22 having trouble with the physical fitness of
23 people coming into the academy because in fact,
24 they had taken the physical, the PPT, shall we
25 say, three years earlier, possibly, and in that

1 Cline

2 time, they had become not physically fit.

3 Q. Okay.

4 A. So it wasn't a current measure of

5 their physical fitness. And they were having

6 problems with people who entered the academy who

7 were not able to go through basic training, even

8 though they had passed the PPT two or three years

9 ago.

10 Q. Okay.

11 A. And they were having a hard time

12 figuring out what to do with it. So when I

13 talked to them, one of the things they said to me

14 was that they would much rather have the CPAT as

15 an entry-level sort of immediately entry-level

16 screen to the academy than to have the PPT,

17 because then at least they would have a current

18 measure of the candidate's physical fitness.

19 And since we knew that the CPAT had

20 in fact been validated, and in fact that the EEOC

21 had used it and had entered into an agreement, we

22 thought that it made sense and the Fire Academy

23 people thought it made sense to use the CPAT as

24 an immediate measure of a candidate's physical

25 fitness rather than to take one at the time of

1 Cline

2 they had standardized the physical exam scores
3 and the written exam scores, and weighted them
4 fifty-fifty.

5 So I did some -- I looked at some of
6 that data. And -- and so I knew that, you know,
7 that they had standardized and weighted them
8 fifty-fifty, and I believe it was probably all
9 three of them, 2043, 7029, and probably 0084,
10 though I can't swear to any of that 'cause all I
11 know is I looked at one, of the data from one of
12 those tests.

13 Q. And in what way do you understand
14 that the written exam and the PPT carried equal
15 weight?

16 A. I mean that when they were
17 standardized, they were basically, the standard
18 scores were averaged or you could say that each
19 part, each standard score was multiplied by .5.

20 Q. In your review of the weighting of
21 the written exam and PPT for exam 7029 or 2043,
22 did you note any challenges or difficulties
23 associated with equally weighting the written and
24 PPT exams?

25 A. Well, one of the things that I noted

1 Cline

2 was that they didn't have scores for people who
3 failed the PPT. They only gave them, like, I
4 think a score of 60 or something. So they didn't
5 have complete score data.

6 Q. Why was that a problem or difficulty?

7 A. It makes it difficult to calculate a
8 standard score.

9 Q. Why is that?

10 A. Because you don't have the entire
11 score distribution, because everyone who was
12 failing was just given a single score.

13 Q. Okay. Did you ever perform any
14 analysis regarding the test score distribution
15 for the written exam and PPT underlying exam 7029
16 and 2043?

17 A. I did. But I'm not sure exactly what
18 I noted except for the fact that we didn't have a
19 complete score distribution for the physical --
20 well, you know, and the other thing I guess is
21 because you had to take -- pass the written to
22 take the physical. So that also creates some
23 problems because you're curtailing the score
24 distribution. You're not giving it to the same
25 population because a larger population takes the

1 Cline

2 written than takes the physical.

3 Q. What did you conclude from that, or

4 what is the difficulty with that?

5 A. It just makes it difficult to

6 standardize.

7 Q. Okay.

8 THE WITNESS: Excuse me, I need to

9 take a break.

10 MS. SEELEY: Sure.

11 THE WITNESS: Yes.

12 (Recess taken.)

13 EXAMINATION (Cont'd.)

14 BY MS. GELLER:

15 Q. Just staying with a discussion of the

16 scoring and equal weighting of 7029 and 2043, did

17 you note that there were only three possible

18 passing scores on the PPT, so distribution is not

19 smooth?

20 A. I did.

21 Q. And does that create, does this

22 create any challenges in terms of combining the

23 written and the physical scores, the PPT scores?

24 A. It might. You know, I'm just not

25 even sure how those scores were derived. I just

1 Cline

2 don't know.

3 Q. You just noted it.

4 A. I just noted that there were only,

5 yes, three, and then everybody who had the

6 failing score was given, I think, a 60. And I

7 believe there were things where some people

8 weren't even allowed possibly to complete the

9 test because they had failed certain events or

10 were having physical difficulty or -- so there

11 were -- there were things there that would make

12 it difficult to do psychometrically as well.

13 Q. Now, did you note that the

14 distributions for the PPT, the score

15 distributions for the PPT and the written exam,

16 were bunched together at the high end?

17 A. I don't remember that. They may well

18 have been. I think so. But I don't remember it

19 myself.

20 Q. If that were true, that the score

21 distributions for the written exam and the PPT

22 were bunched together at the high end, would this

23 create any challenges for combining the written

24 and the physical scores?

25 A. Not necessarily.

1 Cline

2 Q. 7029.

3 A. -- right.

4 Q. So when was the task list updated?

5 Was it updated for the job analysis for exam

6 7029?

7 A. Am I 7029? No. It probably was. I

8 don't really know what was done on either 2049 or

9 7029. What I do know was that I updated the task

10 list for --

11 Q. 6019?

12 A. -- 6019.

13 Q. So you know that you updated the task

14 list for 6019. But you do not know whether the

15 task list was updated for 7029 or for 2043?

16 A. Well, I think if 7029 was Matt's,

17 then he updated it. I'm pretty sure that he

18 wrote that he did that in the job analysis.

19 Q. Okay. Now, do you have any

20 information that the job of entry-level

21 firefighter changed substantially between the

22 time exam 0084, that's the Landy Jacobs that you

23 were involved with, was developed, and exam 7029,

24 that's the Matt Morrongiello exam, was developed?

25 A. The only thing I can speak to is, it

1 Cline

2 didn't seem to me that the job had changed much
3 from when I did observations for the Landy Jacobs
4 job analysis.

5 Q. Okay.

6 A. Which was probably somewhere around
7 1990.

8 Q. Okay.

9 A. I'm not sure. And when I talked to
10 people and did observations in 2007, it seemed to
11 me that the job was pretty much the same. And I
12 also did actually probe them. When we were doing
13 task reviews for the job analysis for this latest
14 job analysis, I probed them about how the job had
15 changed or whether it might have changed after
16 the World Trade Center, after 9/11, because I had
17 found with, for example, Port Authority Police
18 lieutenant or sergeants, that their jobs had
19 changed considerably as a result of that.

20 So I did probe them some on that as
21 part of the job analysis.

22 Q. Okay.

23 A. And they didn't, they really couldn't
24 give me very much in terms of additional things
25 they were doing differently at the firefighter

1 Cline

2 level. I'm sure they do it at the higher levels.

3 Q. Now, in terms of updating the task
4 list, we talked about updating the task list for
5 6019. You state in your report on page 3 that
6 the most recent task list was identified to be
7 used as an initial draft for the current effort.

8 A. Right.

9 Q. What do you mean by "the most recent
10 task list"?

11 A. Well, I thought, at the time, I was
12 talking about Mr. Morrongiello's task list.

13 Q. Okay.

14 A. I didn't realize that in fact there
15 had been another task list update or something in
16 between Mr. Morrongiello and mine. I really
17 didn't realize that until I was talking with
18 Mr. Frankel and Mr. Sample in this preparation.

19 Q. So you were, when you wrote that the
20 most recent task list had to be updated, you were
21 referring to the task, the final task list from
22 exam 7029, correct?

23 A. I thought I was. But they gave me
24 the most recent task list, so it might have been
25 Mr. Johnston's.

1 Cline
2 was determined that it was appropriate to use
3 them as the starting point for a new Job Analysis
4 Questionnaire. However, it was felt that the
5 abilities list might be updated, explicated and
6 made more comprehensive to target more than
7 physical and cognitive aspects of the job; e.g.,
8 interpersonal abilities."

9 Was the previous job analysis
10 referenced in the statement the one conducted for
11 exam 7029?

12 A. Yes, in and in fact, I think they had
13 the same, pretty much the same ability list,
14 pretty much throughout, because it's Fleishman's
15 taxonomy.

16 Q. You mean the 18 Fleischman cognitive
17 abilities included in the 7029 Job Analysis
18 Questionnaire?

19 A. Yes.

20 Q. I have to ask it. Why did you feel
21 that the abilities list used in the previous job
22 analysis might be updated, explicated and made
23 more comprehensive?

24 A. Well, there's a movement in the field
25 of selection and testing that one way to mitigate

1 Cline

2 some of the impact that typically occurs in
3 cognitive tests against protected classes is to
4 include more attributes or abilities than
5 cognitive abilities in the predictor instrument;
6 i.e., in the written test or in the testing
7 mechanism, selection mechanism.

8 So this is cognizant with the
9 recognition that, in the last, say, ten or
10 fifteen years, there's been a movement in
11 selection to use more personality instruments as
12 part of the selection mechanism, as a way of
13 possibly reducing impact.

14 Q. Okay.

15 A. So that's been a movement in testing
16 and selection for the last, I'd say, it's coming
17 to the fore in the last ten years. And so both
18 Dr. Bobko and I were aware of this, and so we
19 wanted to expand what we put on the written test
20 in the hope of reducing some of the impact
21 against protected classes.

22 Q. Okay. Now, did the ability list for
23 the 7029 job analysis include physical ability?

24 A. I don't know. It might have.

25 Q. Did it include interpersonal ability?

1 Cline

2 many different jobs and the abilities they
3 required. And that was the predecessor to O*NET.
4 And I had certainly looked at those taxonomies.

5 Q. When was the predecessor taxonomy of
6 O*NET available?

7 A. It was certainly available in the
8 '80s and the '90s. The O*NET, I think
9 probably -- I'm not sure when the transition was
10 made.

11 Q. Okay. So the Fleischman and -- I'm
12 not sure how to pronounce --

13 A. Quaintance.

14 Q. -- Quaintance's taxonomy was
15 developed in 1984, is that correct?

16 A. Yes.

17 Q. And the Thornton and Byham taxonomy
18 was developed in 1982, is that correct?

19 A. Yes.

20 Q. And you said you're not sure when the
21 O*NET taxonomy became available, but it's
22 basically a successor to a similar one that was
23 available in the '80s.

24 A. Right.

25 Q. And did the predecessor to O*NET also

1 Cline

2 abilities in a written format?

3 A. When you say "any of the above,"

4 you're talking about specific --

5 Q. Oral.

6 A. -- oral?

7 Q. Comprehension, speed of closure,

8 flexibility of closure, perceptual speed or time

9 sharing.

10 A. Okay. Oral comprehension, I

11 definitely worked on tests which presented, you

12 know, a tape recording or a video of speech or

13 actions, but required -- required listening or

14 oral comprehension in order to answer a question,

15 which was then written in a multiple-choice

16 written format.

17 So the stimulus would have been a

18 recording, some sort of recording, and that would

19 be -- so that you would hear speech so that then

20 you could assess it in a written test.

21 Q. Do you know for what positions you

22 developed exams that tested oral comprehension?

23 A. Well, the only time I -- certainly, I

24 did it all the time on oral examinations, okay?

25 I'm thinking in terms of written examinations. I

1 Cline

2 did some of that in some of the practical
3 examinations that I did in my environmental
4 health studies many years ago.

5 Q. Do you know when these exams were
6 developed?

7 A. Oh, these would have been back in the
8 early '80s.

9 Q. Okay.

10 A. When I worked at Psychological
11 Corporation which would have been in the early
12 '70s, we did something called the oral directions
13 test. And I actually did work on that test and a
14 computerized version of that test in which you
15 would give a person instructions orally via
16 recording or via a computer simulation or via --
17 actually, it wasn't a computer, it was a video
18 which was new enough back then, it was a video
19 and the person would tell the person to draw a
20 circle or to put an X in the box or do whatever.

21 And they would have to do whatever,
22 sort of, on the test sheet. So I've worked,
23 those are two off the top of my head I can
24 remember doing.

25 Q. How about speed of closure on a

1 Cline

2 written test?

3 A. Speed of closure, I've never measured
4 before this time. Speed of closure, as I think I
5 said somewhere else, or maybe not, there aren't
6 that many tests of speed of closure and the ones
7 they have are really abstract.

8 Q. How did you know speed of closure
9 could be tested on a written test?

10 A. Because if you research speed of
11 closure, you do find that in fact, it has been
12 measured on written tests in picture formats.

13 Q. Do you know when you had seen speed
14 of closure tested in written formats?

15 A. I looked it up when I was considering
16 this. I think I saw some examples in a basic
17 testing book, Anne Anastasi's Essentials of
18 Psychological Testing, or something like that.
19 And it had some examples of this.

20 Q. What about flexibility of closure?

21 A. That was a harder one but I think --
22 I know I found examples when we were looking it
23 up.

24 Q. Have you ever --

25 A. No, I've never done it myself. This

1

2 UNITED STATES DISTRICT COURT

3 EASTERN DISTRICT OF NEW YORK

4 -----x

THE UNITED STATES OF AMERICA,

5

Plaintiff,

6

and

7

VULCAN SOCIETY, INC., for itself and on
8 behalf of its members, CANDIDO NUNEZ,
MARCUS HAYWOOD and on behalf of a Class
9 of All Others Similarly Situated,

10 Plaintiff-Intervenors,

11 -a-

12 CITY OF NEW YORK, FIRE DEPARTMENT OF
THE CITY OF NEW YORK, NEW YORK CITY
13 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
SERVICES, MAYOR MICHAEL BLOOMBERG and
14 NEW YORK CITY FIRE COMMISSIONER
NICHOLAS SCOPETTA, in their Individual
15 and Official capacities,

16

Defendants.

17 -----x

18 271 Cadman Plaza
Brooklyn, New York

19

March 24, 2008
20 9:51 a.m.

21

22

23 DEPOSITION OF CATHERINE CLINE

24

25

1 CATHERINE CLINE

2 sense to combine them. So I had him
3 correct the correlations for the
4 reliability of each variable that
5 entered into the correlation.

6 Q. What did you find out?

7 A. I can't remember. I mean, I
8 remember that, in fact, the
9 correlations went up pretty high, but
10 they usually do on cognitive tests.
11 You usually have a fairly good -- the
12 test scores do correlate pretty highly.
13 Not totally. There are still some
14 distinct variances, but they do
15 correlate pretty highly.

16 Q. Now, if an item is measuring
17 the ability it's intended to measure,
18 that is, items that measure the same
19 ability should be more highly
20 correlated with each other than with
21 items measuring other abilities; is
22 that correct?

23 A. Yes.

24 Q. What would it indicate to you
25 if items intended to measure one

1 CATHERINE CLINE

2 A. No.

3 Q. Why not?

4 A. Because my computer wouldn't
5 do it.

6 Q. Okay. So if you had
7 performed an analysis of the pattern of
8 correlations between individual test
9 items, wouldn't this have told you
10 whether an item is measuring the
11 ability it's intended to measure, that
12 is, items that measure the same ability
13 should be more highly correlated with
14 each other than with items measuring
15 other abilities?

16 A. Yes. I would think so, yes.

17 Q. And if items measure the
18 abilities or attributes they're
19 supposed to, you would expect to see
20 that pattern correlation; is that
21 right?

22 A. Yes.

23 Q. What would it indicate to you
24 if items intended to measure one
25 cognitive ability correlate as or more

1 CATHERINE CLINE

2 A. I think not. I think it was
3 just something that Dr. Bobko and I
4 discussed and that was why we wanted to
5 add in more abilities or more personal
6 characteristics or whatever you wanted
7 to call them into the test mix because
8 we knew from the research that some of
9 these types of characteristics, that
10 measures of them had less impact
11 against minority groups than, say,
12 straight cognitive tests did. So we
13 wanted to add in more of these types of
14 abilities so that we could sort of
15 increase the mix of abilities and
16 possibly come up with a more
17 comprehensive test, but a test which
18 also had less impact on minorities.

19 Q. And nobody from DCAS spoke to
20 you about that?

21 A. No. This is more the way the
22 field is going in terms of -- that's
23 the way the field is going in terms of
24 reducing adverse impact, the inclusion
25 of more different and varied predictor

508

1 CATHERINE CLINE

2 measures in the examination.

3 And so Phil and I knew that,

4 but I don't think DCAS people would

5 have necessarily known that, and maybe

6 not even the lawyers would have known

7 that. But it certainly -- it's

8 certainly -- and certainly the Fire

9 Department wouldn't have known it.

10 But it's certainly something

11 that is of interest in my field. And

12 in the last 10 years there's been a lot

13 of information that's come out that, in

14 fact, yes, situational judgment tests

15 are valid. They do have valid

16 validity. There's been meta analyses

17 conducted that shows that the

18 situational judgment exercises do have

19 validity, that they can be extended to

20 other construct besides judgment.

21 There's been a stream of research that

22 personality measures, when added to

23 cognitive measures, also increase the

24 validity while decreasing impact with

25 minority groups.

1 CATHERINE CLINE

2 So a lot of this research has
3 been in the last 10 years. So while
4 Dr. Bobko and I were certainly aware of
5 it, it's not necessarily something that
6 people at DCAS would have been aware
7 of. So we were of the opinion that
8 that would be one way to go to decrease
9 the impact because that was a concern
10 for both of us.

11 Q. Isn't it a fact that the
12 situational judgment measures and
13 certainly the personality measures go
14 back for 20 years or more?

15 A. Actually, they haven't been
16 used in an employment setting for that
17 long. It's been used more for
18 counseling. It's only been in the last
19 10 or 15 years that people have used
20 the personality measures as ways of
21 predicting job performance, and that
22 there's been data that shows that, in
23 fact, they're valid predictors of job
24 performance in a lot of cases.

25 But there's also been a lot

513

1 CATHERINE CLINE

2 Q. And New York City is the
3 biggest city in the country?

4 A. I think so.

5 Q. And the Fire Department has
6 had roughly three percent

7 African-American incumbents for the
8 last 20 or 30 years?

9 A. And even fewer females.

10 Q. And even fewer females?

11 A. Yes.

12 Q. And when you worked at DCAS,
13 it was -- the years, was it '84 to '90?

14 A. Something like that.

15 Q. And those statistics that I
16 just -- we just talked about, they were
17 standard at that time as well?

18 A. Yes.

19 Q. And at that time, people also
20 knew about personality measures?

21 A. No. That really -- I'll tell
22 you that really I started becoming more
23 conversant with the use of personality
24 measures a couple of years after I left
25 the City, when I attended a FIAT

514

1 CATHERINE CLINE

2 convention and the huge talk at that

3 FIAT convention was what they call the

4 army project alpha or something. And

5 this was a large army project which

6 looked at personality tests as

7 independent indicators of job

8 performance in the army. And so that

9 would have been in 1993 or so. And

10 that was the first time people started

11 to take personality tests seriously as

12 a predictor of job performance.

13 Because before that people

14 were really -- people can see what

15 they're supposed to be answering, they

16 can see what's socially desirable here

17 to say. They can take these tests. So

18 they were really not considered in good

19 repute for a long time. But with the

20 presentation of the data from the army

21 project, they became much more credible

22 and people started looking at ways to

23 incorporate personality testing into

24 employment testing.

25 Q. Did the people include the

EXHIBIT CC

OBJECTION AND RESPONSE TO ADMISSION NO. 8:

Denied.

ADMISSION NO. 9:

Defendants neither conducted nor arranged for anyone else to conduct an adverse impact study regarding the impact of Written Exam 2043 on black candidates prior to the commencement of litigation in this case.

OBJECTION AND RESPONSE TO ADMISSION NO. 9:

Denied.

ADMISSION NO. 10:

The FDNY was aware at least as of early 2000 that the results of Written Exam 7029 violated the 4/5ths or 80% rule in the EEOC's *Uniform Guidelines on Employee Selection Procedures* with respect to black candidates. (See Letter dated May 25, 2000 from Frank R. Nicolazzi to Commissioner Von Essen; EEPD 0533.)

OBJECTION AND RESPONSE TO ADMISSION NO. 10:

Denied, but otherwise admit that at least as of early 2000, the FDNY was aware that the EEPD took the position that the results of Written Exam 7029 violated the 4/5ths or 80% rule with respect to black candidates. (See Letter dated May 25, 2000 from Frank R. Nicolazzi to Commissioner Von Essen; EEPD 0533.).

ADMISSION NO. 11:

DCAS used the same job analysis and test development report for the creation of Written Exam 2043 as had been used for Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 11:

Admit.

OBJECTION AND RESPONSE TO ADMISSION NO. 19:

Admit.

ADMISSION NO. 20:

As of 1999, defendants were aware that:

- a. the percentage of black firefighters in New York City was approximately 2.90%;
- b. the percentage of black firefighters in Los Angeles, CA was 14.0%;
- c. the percentage of black firefighters in Chicago, IL was 20.4%;
- d. the percentage of black firefighters in Houston, TX was 17.1%;
- e. the percentage of black firefighters in Philadelphia, PA was 26.3%;
- f. the percentage of black firefighters in San Diego, CA was 7.7%;
- g. the percentage of black firefighters in Dallas, TX was 18.1%;
- h. the percentage of black firefighters in San Antonio, TX was 7.0%; and
- i. the percentage of black firefighters in San Jose, CA was 7.1%;

(See Appendix III to Letter from Frank Nicolazzi, Vice Chairman of EEPC, to Commissioner Von Essen dated May 25, 2000; EEPC 0540.)

OBJECTION AND RESPONSE TO ADMISSION NO. 20:

Admit as to subpart a. and denied as to subparts b. through i. In addition, as of 1999, defendants were aware that the EEPC took the position that the percentages of black firefighters in the cities represented by subparts b. through i. were as indicated. (See Appendix III to Letter from Frank Nicolazzi, Vice Chairman of EEPC, to Commissioner Von Essen dated May 25, 2000; EEPC 0540.)

ADMISSION NO. 21:

As of 2000, defendants were aware that:

- e. the percentage of black firefighters in Los Angeles, CA was 15.8%, with 2.3% black new hires;

(See Determination in EEOC Charge# 160 2002-0128, USA 000028 – USA 000033.)

OBJECTION AND RESPONSE TO ADMISSION NO. 22:

Denied.

ADMISSION NO. 23:

Defendants have neither conducted nor arranged for anyone else to conduct an adverse impact study regarding the impact on black candidates of the college credit requirement for appointment to the position of entry-level firefighter.

OBJECTION AND RESPONSE TO ADMISSION NO. 23:

Admit.

ADMISSION NO. 24:

Defendants neither conducted nor arranged for anyone else to conduct an adverse impact study regarding the impact on black candidates of the college credit requirement for appointment to the position of entry-level firefighter prior to the commencement of litigation in this case.

OBJECTION AND RESPONSE TO ADMISSION NO. 24:

Admit.

ADMISSION NO. 25:

Defendants have neither conducted nor arranged for anyone else to conduct an adverse impact study regarding the impact on black candidates of the driver's license requirement for appointment to the position of entry-level firefighter.

INTERROGATORIES

INTERROGATORY NO. 33:

For each Request for Admission contained in Plaintiffs-Intervenors' First Set of Requests for Admission (Nos. 1-43) which the defendants deny in whole or in part, state all facts and identify all documents on which the defendants' denial is based, and further, identify all persons with knowledge of the facts stated or the location of the documents identified.

OBJECTION AND RESPONSE TO INTERROGATORY NO. 33:

Defendants object to this request as plaintiff-intervenors have exceeded the amount of interrogatories allowed under the Federal Rules of Civil Procedure and the parties' Amended Scheduling Order. In addition, defendants object to this request to the extent it seeks disclosure of attorney work-product and attorney-client communications.

Dated: New York, New York
October 20, 2008

MICHAEL A. CARDOZO
Corporation Counsel of the City of New York
Attorney for Defendant
100 Church Street, Room 2-100
New York, New York 10007
(212) 788-1247
gpestana@law.nyc.gov

By: 

GEORGIA PESTANA
Assistant Corporation Counsel

OBJECTION AND RESPONSE TO ADMISSION NO. 60

In so far as Exam 2043 was based in part on Dr. Landy's work for Exam 0084, the request is denied, but otherwise admit that no psychometrician was employed or engaged solely to develop Exam 2043.

ADMISSION NO. 61

DCAS Examiner Matthew Morrongiello did not conduct a content validity study with respect to Exam 7029. (See Morrongiello Deposition January 14, 2008 at Tr. 75:9-11).

OBJECTION AND RESPONSE TO ADMISSION NO. 61

Admit.

ADMISSION NO. 62

No one has conducted a construct validity study with respect to Exams 7029 or 2043 that conformed with the requirements of the *Uniform Guidelines*.

OBJECTION AND RESPONSE TO ADMISSION NO. 62

Defendants object to the request to the extent the request implies that the *Guidelines* require that a construct validity study be performed, but otherwise admit that a construct validity study was not performed with respect to Exams 7029 and 2043.

ADMISSION NO. 63

All other things being equal, an examination that measures more of the knowledge, skills, abilities or characteristics ("KSACs") that are important for a job is expected to be more valid than an examination that measures fewer of the KSACs that are important for that job.

OBJECTION AND RESPONSE TO ADMISSION NO. 63

Admit.

ADMISSION NO. 64

Each of the following KSACs is non-cognitive: (Please admit or deny as to each KSAC.)

- a. Resistance to stress;
- b. Teamwork;
- c. Responsibility;
- d. Desire to learn;
- e. Honesty;
- f. Cleanliness;
- g. Medical interest;
- h. Following instructions;
- i. Achievement orientation;
- j. Dependability;
- k. Conscientiousness; and
- l. Process monitoring.

OBJECTION AND RESPONSE TO ADMISSION NO. 64

Denied as to subparts h. and l., but otherwise admit as to subparts a. through g. and i. through k.

ADMISSION NO. 65

Each of the KSACs listed in paragraph "64" is important to the job of firefighter. (See Schemmer deposition July 8, 2008 at Tr. 294). (Please admit or deny as to each KSAC.)

OBJECTION AND RESPONSE TO ADMISSION NO. 65

To the extent the request seeks an admission that each of the KSAC's listed in paragraph "64" are, together as a group, important to the job of an entry-level firefighter in the F.D.N.Y. the request is denied, but otherwise admit that disjunctively, each of the KSAC's listed in paragraph "64" have been, on occasion, found to be important to the job of firefighter.

ADMISSION NO. 66

Each of the KSACs listed in paragraph "64" could have been tested for using a written exam format at the time that Written Exams 7029 and 7043 were administered. (See Schermer deposition July 8, 2008 at Tr. 294). (Please admit or deny as to each KSAC.)

OBJECTION AND RESPONSE TO ADMISSION NO. 66

Denied.

ADMISSION NO. 67

Had the KSACs listed in paragraph "64" been tested for on Written Exams 7029 and 2043, using well developed items, and weighting the items according to the importance of the KSACs, the testing of each would have been expected to add to the validity of Exams 7029 and 2043. (Please admit or deny as to each KSAC.)

OBJECTION AND RESPONSE TO ADMISSION NO. 67

Defendants object to the request as vague and ambiguous. Notwithstanding and without waiving the foregoing objection, the request is denied.

ADMISSION NO. 68

Had the KSACs listed in paragraph "64" been tested for on Written Exams 7029 and 2043, using well developed items, and weighting the items according to the importance of

the KSACs, Exams 7029 and 2043 would have been expected to have less adverse impact against black test-takers. (Please admit or deny as to each KSAC.)

OBJECTION AND RESPONSE TO ADMISSION NO. 68

Defendants object to the request as vague and ambiguous. Notwithstanding and without waiving the foregoing objection, the request is denied.

ADMISSION NO. 69

All of the non-cognitive abilities tested for on Exam 6019 are KSACs for which written tests were available prior to 1999.

OBJECTION AND RESPONSE TO ADMISSION NO. 69

Denied.

ADMISSION NO. 70

Written Exam 6019 was at least as job-related as Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 70

Admit.

ADMISSION NO. 71

Written Exam 6019 was at least as job-related as Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 71

Admit.

ADMISSION NO. 72

Written Exam 6019 was at least as consistent with business necessity as Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 72

Admit.

ADMISSION NO. 73

Written Exam 6019 was at least as consistent with business necessity as Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 73

Admit.

ADMISSION NO. 74

Written Exam 6019 had less adverse impact against black test takers with respect to their passing rates than Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 74

To the extent the request seeks an admission that Written Exam 7029 had an adverse impact against black test takers the request is denied, but otherwise admit that a greater number of black candidates passed Written Exam 6019 than Written Exam 7029.

ADMISSION NO. 75

Written Exam 6019 had less adverse impact against black test takers with respect to their passing rates than Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 75

To the extent the request seeks an admission that Written Exam 2043 had an adverse impact against black test takers the request is denied, but otherwise admit that a greater number of black candidates passed Written Exam 6019 than Written Exam 2043.

ADMISSION NO. 76

Written Exam 6019 had less adverse impact against black test takers with respect to their rank ordering than Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 76

To the extent the request seeks an admission that Written Exam 7029 had an adverse impact against black test takers the request is denied, but otherwise admit that a greater number of black candidates achieved a higher ranking on Written Exam 6019 than Written Exam 7029.

ADMISSION NO. 77

Written Exam 6019 had less adverse impact against black test takers with respect to their rank ordering than Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 77

To the extent the request seeks an admission that Written Exam 2043 had an adverse impact against black test takers the request is denied, but otherwise admit that a greater number of black candidates achieved a higher ranking on Written Exam 6019 than Written Exam 2043.

ADMISSION NO. 78

The "situational judgment exercise" items developed for Written Exam 6019 could have been developed and used on Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 78

Denied.

ADMISSION NO. 79

The "situational judgment exercise" items developed for Written Exam 6019 could have been developed and used on Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 79

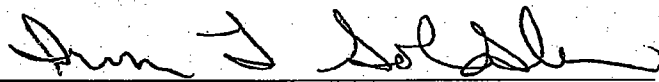
Denied.

EXHIBIT DD

Report in the matter of

United States of America v. City of New York

By

A handwritten signature in black ink, appearing to read "Irwin L. Goldstein", written over a horizontal line.

Irwin L. Goldstein, Ph.D.
President, Organizational and Personnel Research, Inc.
Rockville, Maryland

July 31, 2008

An item-writing panel consisting of one Lieutenant and four firefighters wrote the test questions (or "items"). (*Id.* at pp. 11-12). The test questions were then reviewed by a panel of one FDNY Lieutenant and four firefighters. (*Id.* at p. 12).

The DCAS Examiner who developed Written Exam 2043, Alberto Johnston, did not conduct a new job analysis. He did not administer a job analysis survey, did not convene a new linking panel and did not develop a new test plan. Instead, he used the test plan developed for Written Exam 7029. Another item-writing panel, consisting of four firefighters and one Lieutenant, wrote new test questions for Written Exam 2043. (Johnston Dep. at pp. 23-28, 249; Pl. Dep. Exh. 104).

SECTION IV.

ANALYSES OF PROFESSIONAL STANDARDS AND THE CITY'S USE OF WRITTEN EXAMS 7029 AND 2043

A. THERE ARE SEVERAL VALIDATION APPROACHES RECOGNIZED BY PROFESSIONAL STANDARDS, INCLUDING CONTENT, CRITERION-RELATED AND CONSTRUCT APPROACHES.

The Uniform Guidelines discuss several approaches to validating a selection procedure, or gathering an appropriate set of evidence establishing that the selection procedure is job related.⁶ One approach discussed in the Uniform Guidelines is "content" validity.⁷ To establish content validity, it is necessary to show that the test is representative of significant portions of the job. In this case, it is necessary to provide evidence establishing that the test samples (or requires the use of) behaviors that are critical for job performance. Note that, with content validity, no measure of job performance is collected. Rather, it is assumed that, if the critical behaviors being measured on the test overlap the critical behaviors on the job, then it is likely that performance on the test will

⁶ It should be noted that a selection procedure need not be a written test like those at issue here. Selection procedures also may include such things as minimum qualifications (e.g., a high school diploma requirement), an oral interview or a job simulation exercise. Nevertheless, for simplicity, I will refer to selection procedures as "tests" in this report.

⁷ Other approaches discussed in the Uniform Guidelines are known as "criterion-related" (or "predictive" validity) and "construct" validity. Criterion-related validity requires the establishment of a statistical relationship between performance on the test and performance on the job. "Construct" validity involves establishing that a test measures a particular construct or trait (such as leadership ability, spatial ability or deductive reasoning) and then showing that the construct measured by the test is related to job performance. Over the years, most construct validation studies have involved conducting criterion-related analyses or establishing that a test designed to measure a particular construct correlated well with other tests known to measure the same construct, and that criterion-related validity already has been established for the other tests. As stated previously, I understand that other expert reports may address criterion-related and construct validity. This report provides my opinions regarding content validity.

reflect performance on the job. Because actual performance on the job is not being measured when a content validity methodology (as opposed to a criterion-related methodology) is used to validate a test, the use of a robust job analysis which carefully identifies the critical behaviors to be measured is essential. It is only by using data from such a job analysis that a test is likely to be developed which will reflect the behaviors that are important on the job.⁸ In addition, as part of establishing content validity, it is critical to use Subject Matter Experts ("SMEs") to independently make judgments of whether the test items developed overlap the important work behaviors, activities and/or worker Knowledge, Skills and Abilities and other characteristics ("KSAOs") identified by the job analysis. Thus, content validity, as compared to criterion-related validity, is especially dependent upon a well-done job analysis, and careful procedures are necessary in order to have confidence in the judgments about the relationship of the test to the job. Of course, it is not the case that job analysis is unimportant when using a criterion-related validity strategy; the job analysis helps establish what type of test should be used and what type of job performance ("criterion") should be used. However, if mistakes are made in the job analysis, the statistical relationship (correlation) between the test and job performance will not be established or will be weak. There is no such statistical relationship for content validity. Thus, for a content validity study, the job analysis and carefully established judgment procedures become especially critical.

As noted in the Principles for the Validation and Use of Personnel Selection Procedures (SIOP (2003), p. 4), the conceptualization of validity over the years has evolved into treating validity as a "unitary" concept, and it is necessary to use the most appropriate strategies to develop evidence of job relatedness. The objective is always to use appropriate methods to establish lines of evidence which support the use of the test at issue.

B. A CONTENT VALIDITY APPROACH IS NOT APPROPRIATE FOR ESTABLISHING THE VALIDITY OF WRITTEN EXAMS 7029 AND 2043.

There are situations where each of the three validation strategies explained above is not appropriate. For example, there are situations where it is difficult to obtain reliable and appropriate job performance data and, in that situation, it might not be possible to use a criterion-related validity approach. On the other hand, if the test to be validated measures a broad construct, such as leadership ability, spatial ability or deductive reasoning, content validity is not appropriate because it is difficult to establish that the critical behaviors being measured on the test overlap the critical behaviors on the job.

Although the City's experts know of no reason it would not be possible to do so (Bobko Dep., p. 234; Schemmer Dep., p. 57), the City has not conducted a criterion-related

⁸ In addition, the Uniform Guidelines note that content validity is not an appropriate strategy when a selection procedure involves knowledge, skills and abilities which an employee will be expected to learn on the job (29 C.F.R. 1607(C)(1)). As discussed in the following section of this report, the job analysis approach DCAS used for Exam 7029 did not consider or measure this factor.

Pl. Dep. Exh. 149, p. 12). However, on review of Inductive Reasoning questions originally written for use in Written Exam 6019, Dr. Bobko apparently stated that the questions did not measure "the kind of inductive reasoning [firefighters] talked about at site visits" attended by Dr. Bobko. (DCAS 0013909). In order to establish content validity, it is necessary to show that the test questions actually mirror the content of the job.

No such evidence is presented by the City. The Morrongiello Report ends with the development of the items for Written Exam 7029 and does not report any attempt to link them back to the important aspects of the job. Thus, the DCAS Examiner responsible for the development of Written Exam 7029 was correct when, as noted above, he stated in his deposition that he has never conducted a content validity study and that his test development report is not a validity report. (Morrongiello Dep., p. 75). Similarly, the DCAS Examiner responsible for developing Written Exam 2043 (using the Written Exam 7029 test plan) correctly indicated in his deposition that he also has never conducted a content validity study. (Johnston Dep., p. 76). Without the crucial step of linking the content of the test (the test questions) back to the content of the job, a claim of content validity cannot be justified.

DCAS did not conduct such analyses for either Written Exam 7029 or Written Exam 2043. To the extent that the City might claim that the review of Written Exam 7029 and 2043 items by the Examiners who developed each of the written examinations and by others in DCAS is sufficient to establish the necessary linkage between test content and job content, the City would be incorrect. The procedures followed are not the kind of systematic validity exercise needed. As stated previously, a new written firefighter examination, Written Exam 6019, has been developed by Dr. Cline and administered by the City. Dr. Cline attempted to conduct a content validity analysis, asking five SMEs, who were DCAS Test and Measurement Specialists (i.e., Examiners) and had experience writing cognitive test items, to independently judge whether each test item was a measure of the ability it was originally designed to measure. (6019 Test Development Report, Pl. Dep. Exh. 200, p. 7). Dr. Cline concluded that DCAS Examiners she used to review the Written Exam 6019 items had difficulty differentiating between the various cognitive abilities that the written examinations were supposed to measure. (Cline Dep., pp. 281-90). Specifically, she found that the DCAS Examiners were able to match only "62.5% of the 40 cognitive items on Written Exam 6019 to the intended classifications." (6019 Test Development Report, Pl. Dep. Exh. 200, p. 7). She further found that the DCAS Examiners had a problem distinguishing between items intended to measure Inductive Reasoning and Deductive Reasoning and also between Spatial Orientation and Visualization. (Id.).

As noted in the Principles (SIOP, 2003), "evidence for validity based on content rests on demonstrating that the selection procedure adequately samples and is linked to the important work behaviors, activities, and/or worker KSAOs defined by the analysis of work" (p. 22). No such evidence has been presented for Exam 7029 or Exam 2043. Thus, content validation evidence for Written Exams 7029 and 2043 was not established.

- b. The City Has Not Established That the Important Work Behaviors and Activities and/or the Important Knowledge, Skills, Abilities and Other Characteristics That Written Exams 7029 and 2043 Purport to Measure Are Needed at Entry, as Required to Support a Claim of Content Validity.

A content valid test should measure work behaviors, activities, and/or worker KSAOs that are important for performance of the job and are needed at entry, rather than learned on the job. As discussed in the Uniform Guidelines (1978), content validity is not an appropriate strategy where the selection procedure involves work behaviors, activities, and/or worker KSAOs that are to be learned on the job (29 C.F.R. 1607.14(C)(1)). As stated in the professional literature (Goldstein, Zedeck and Schneider (1993)), to establish content validity, it is "necessary to determine what is needed at entry and what can be learned in training or on the job" (p. 16). Thus, if an employer wishes to use a content validity approach, the job analysis and test development procedures conducted must ask SMEs to independently judge whether the test components reflect the important KSAOs that are needed at entry.

This is another respect in which a job analysis conducted to support a claim of content validity may differ from one used to develop a test that will be validated using a criterion-related approach. As stated previously, criterion-related validity models attempt to establish the statistical relationship (correlation) between performance on a test and performance on a job. Criterion-related validity is concerned with whether test scores predict job performance. In contrast, content validity models are concerned with establishing that the content of a test reflects the content of a job – in other words, that what a candidate must do to perform well on the test corresponds to what a worker must do to perform well on the job. That critical content validity link is broken if what the worker must do to perform well on the job is learned after entering the job (and, thus, after taking the test). Therefore, to establish content validity, it is critical to prove that the test reflects the KSAOs that are important when entering the job.

There are no analyses presented in the Morrongiello Report establishing that any of the tasks and/or abilities required by the test items are needed at entry. Indeed, notes made by the DCAS Examiner responsible for the development of Exam 7029 reflect that he considered using "Day 1" scales for job tasks and abilities in his job analysis questionnaire but it was decided in discussions with his DCAS supervisor that he would not. (Morrongiello Dep., p. 118). When the Examiner was questioned in his deposition as to why Day 1 scales were not used, he could say only that they had not been used in the previous job analyses by Landy, Jacobs for Exam 0084. (*Id.* at p. 119). This does not excuse the City's failure to use a Day 1 scale in its job analysis for Exam 7029. As stated previously, the Landy, Jacobs Report indicates that Landy, Jacobs intended to conduct a criterion-related validity study for Written Exam 0084. The City has presented no evidence that what is measured by the test items is needed Day 1 on the job and cannot establish that the Written Exams 7029 and 2043 have content validity.

Again, this lack of proper content validity evidence can be contrasted with the work the City has done in connection with its new examination, Written Exam 6019. Beginning in approximately April 2006, a new job analysis was conducted for Exam 6019. In the job analysis questionnaire administered as part of that job analysis, the SMEs (firefighters) completing the questionnaire were asked to indicate whether a new firefighter would be expected to perform each of the firefighter job tasks on Day 1 and whether various abilities were needed Day 1, as well as to rate the importance of the tasks and abilities. (6019 Job Analysis Report, Pl. Dep. Ex. 149, App. A, pp. 4-5). In order to be considered an important task/ability needed at entry, the task or ability had to be rated a 4 (corresponding to "very important") on a five-point scale. (*Id.* at p. 9). In addition, 2/3 of the respondents had to rate the ability or task as needed (or expected to be performed if the occasion arose) on Day 1. (*Id.*). These are the kind of rating data that should be collected in order to determine whether a test is job related using a content validity strategy. Notably, the use of these standards resulted in Dr. Cline dropping from her task list for Exam 6019 43 of the 111 tasks that were used in the final task list for Exam 7029.

Moreover, the emphasis on Day 1 tasks and abilities was carried through to the test development process for Written Exam 6019. During the test development process, Dr. Cline trained the item-writers about the need to use Day 1 tasks linked to the particular ability an item was intended to measure. In yet another step, Dr. Cline had a group of another 12 SMEs (firefighters and lieutenants) rate whether each of the items written for Written Exam 6019 was important for entry-level performance. As a result of that analysis, Dr. Cline reported in her test development report that she dropped a number of test items which did not meet the standard of needed Day 1. (6019 Test Development Report, Pl. Dep. Exh. 200, p. 9). No such procedures were used by DCAS in developing Written Exams 7029 and 2043.

c. The Job Analysis Work that Was Conducted by the City Is Flawed and Cannot Support a Claim of Content Validity.

There are a number of other aspects to the job analysis strategy used by the City as part of DCAS's development of Written Exam 7029, as reflected in the Morrongiello Report, that are flawed and make the report and the analyses it reflects an unreliable foundation for a claim of content validity.

i. The criteria used to include tasks and abilities were not sufficient to support a claim of content validity.

First, in the Exam 7029 job analysis, DCAS utilized a four-point alphabetical rating scale for establishing which abilities and which tasks were important. DCAS converted the alphabetical rating scale to a numeric scale for purposes of analysis. The numeric scale was: 4 for "critical"; 3 for "important"; 2 for "somewhat important"; and 1 for "not important" or "not relevant." (Morrongiello Report, Pl. Dep. Exh. 54, p. 10). The importance criterion used as the standard for inclusion in the final task list was 2.5 on the four-point rating scale, which is between "somewhat important" and "important." (*Id.*). In addition, two-thirds of the firefighters who completed the job analysis questionnaire

had to respond that they performed the task. (*Id.*). According to the DCAS Examiner who developed Written Exam 7029, the criteria for tasks were selected largely based upon whether they resulted in retaining a sufficient number of tasks for use in developing the written examination (and the criteria for abilities were selected in part to be parallel to the criteria for tasks), indicating that the focus was on test development and not on ensuring content validity. (Morrongiello Dep., p. 219). To support a content validity strategy, tasks and abilities must meet more stringent criteria for importance as well as meeting a Day 1 standard.

The analyses conducted by Dr. Cline in connection with Written Exam 6019 used more appropriate criteria and identified a different set of important tasks. First, it should be noted that, as part of the job analysis and test development for Written Exam 6019, Dr. Cline attempted to determine whether the FDNY firefighter job had changed since the early 1990s, when she last studied it, and specifically questioned panels of firefighters about any changes since September 11, 2001. (Cline Dep., p. 73). According to Dr. Cline, she found that the job had not changed appreciably. Yet, using more appropriate criteria to identify important Day 1 tasks, Dr. Cline developed a task list different than that developed by Mr. Morrongiello and used in the development of Written Exam 7029. As noted above, in Dr. Cline's job analysis for Exam 6019, she used a criterion of 4 on a five-point scale, which corresponded to a rating of "very important." Thus, her rating required more than a rating of "somewhat important" to "important." In her deposition, Dr. Cline reported that she would never use "somewhat important" as a scale point because she would want anything she tested for to be important. (Cline Dep., pp. 118-19).

The Bobko/Schemmer Report offers an illustration of the differences that using more appropriate criteria, as Dr. Cline did, can make. The report of Dr. Bobko and Dr. Schemmer states that the tasks in the task category of "Size Up" (such as "Determines appropriate entry into structure for assigned position based on structure type, location of fire, etc."), clearly "invoke cognitive processes." (Bobko/Schemmer Report, p. 28).⁹ However, when Dr. Cline completed her content validity analysis, five of the seven tasks listed by Drs. Bobko and Schemmer as being in the "Size Up" category had not met Dr. Cline's criteria that tasks be both important and needed Day 1 on the job. (6019 Job Analysis Report, Pl. Dep. Exh. 149, App. B, Table 4). For example, the "Size Up" task cited in the parenthetical above ("Determines appropriate entry . . .") did not meet the Day 1 criterion. (*Id.*). As noted above, Dr. Cline's analyses indicated that 43 of the 111 tasks used as a foundation for building test items included in Written Exams 7029 and 2043 did not meet her criteria.

The criteria used by the City in the development of Written Exam 7029 allowed the retention of tasks and abilities that have not been judged to be at least "important" and/or

⁹ Obviously, this type of statement does not constitute a content validity study, nor is it proof that the written examinations have content validity. It must also be noted that the report of Dr. Bobko and Dr. Schemmer does not state that the City has conducted a content validity study, or that Written Exams 7029 and 2043 have content validity.

are not needed upon entry. This does not provide a strong enough foundation for the inclusion of tasks and abilities in support of a content validity approach.

- ii. The linking panel exercise conducted for Exam 7029 does not provide reliable information that can support a claim of content validity.

As stated previously, in addition to using a job analysis questionnaire to collect data regarding tasks and abilities, in developing Written Exam 7029 DCAS used a "linking panel" composed of 12 firefighters to rate how important each of a set of cognitive abilities was for the performance of various task clusters (categories of related tasks). The linking panel ratings were 3 for "critical"; 2 for "important"; 1 for "somewhat important"; and 0 for "not relevant." (Morrongiello Report, Pl. Dep. Exh. 54, p. 10). In developing Written Exam 2043, the City gathered no new ratings, simply relying on the flawed analyses conducted for Exam 7029.

The City did not check the reliability of the linking panel ratings or do any systematic analysis of whether they were consistent. (Morrongiello Dep., p. 299). If the City had done so, it would have found that there are a number of anomalies in the ratings. For example, in one instance a linking panel member gave a rating of 3 for "critical" to nearly all combinations of task clusters and abilities. (Pl. Dep. Exh. 180, Rater #9). This included indicating that Written Comprehension is critical to the performance of the task "trimming broken glass from window frames." It is difficult to see how Written Comprehension relates to this psychomotor task. When the DCAS Examiner responsible for Exam 7029 was questioned in his deposition as to whether he reviewed the ratings, he indicated that "I generally look through the packets. Generally, though, not to see if they make sense." (Morrongiello Dep., p. 318). Indeed, the DCAS Examiner's only explanation for such ratings was that, although they were not supposed to, the linking panel members may have been thinking about training to perform the job task or writing about performing the job task afterward instead of actually performing the job task itself. (*Id.* at pp. 315-18). Such ratings indicate that at least some members of the linking panel either did not understand the linking task they were asked to perform or did not perform it conscientiously. Indeed, one of the raters, Rater #1, indicated in her deposition that she did not understand the definitions of Problem Sensitivity, Inductive Reasoning, Visualization and Information Ordering. (Cziko Dep., pp. 57-62). Others, such as Rater #3, indicated that, upon thinking about it, a number of their ratings were incorrect, and/or that they did not know, at the time they completed the linking panel exercise, what concepts like Inductive and Deductive Reasoning were. (Baretta Dep., pp. 49-50, 93-95). Indeed, Rater # 3 stated that he just guessed. (*Id.* at pp. 65, 72, 93-95). An appropriate job analysis uses methods to ensure that raters like the linking panel members understand the task they are to perform, perform it conscientiously and provide ratings that, at the very least, make sense.

D. FROM A CONTENT VALIDITY PERSPECTIVE, THERE IS NO FOUNDATION FOR THE TEST PLAN THE CITY USED FOR WRITTEN EXAMS 7029 AND 2043 OR THE WAY IN WHICH THE CITY USED THE WRITTEN EXAMINATIONS.

Because of the limited scope of the job analysis conducted for Exam 7029, there is no sound basis for the decisions the City made about which abilities Written Exams 7029 and 2043 would measure or the weights to be given to various abilities in developing and using Written Exams 7029 and 2043. The City also has not justified the pass/fail cutoff scores it used on the written examinations or the weights it gave to the written and physical examinations for purposes of ranking candidates on the eligibility lists.

First, there is the question of whether the components of Written Exams 7029 and 2043 are sufficiently comprehensive. The job analysis questionnaire used an abilities list that included only cognitive abilities taken from a taxonomy that was developed by Fleishman and Quaintance. The taxonomy attempted to specify all the various constructs which make up human performance. The DCAS Examiner responsible for developing Written Exam 7029 indicated that he did not consider including non-cognitive abilities specified in the taxonomy when creating the Exam 7029 job analysis questionnaire. (Morrongiello Dep., pp. 129-130). Also, when the DCAS Examiner responsible for developing Written Exam 2043 was asked in his deposition whether he ever considered non-cognitive abilities, he indicated that he had not. (Johnston Dep., p. 104). There is no good scientific or professional basis given for the decision not to conduct a more comprehensive job analysis and test development process. It is, of course, important to note that tests of cognitive abilities often result in adverse impact against protected classes such as African Americans and Hispanics. The literature also indicates there is considerably less adverse impact in the measurement of non-cognitive personal abilities. Thus, the test development plan is not only incomplete for Exams 7029 and 2043, but it is weighted toward components which tend to produce more adverse impact.

The decision to develop Written Exams 7029 and 2043 as measures of only cognitive abilities can be contrasted with the decisions made more recently in developing Written Exam 6019. In her job analysis for Exam 6019, Dr. Cline established linking data for a wide variety of personal abilities, including Adaptability, Tenacity, Integrity, Work Standards, Resilience, Coordination, and Establishing and Maintaining Interpersonal Relationships. (6019 Job Analysis Report, Pl. Dep. Exh. 149, App. B, Table 5). Indeed, Dr. Cline's analysis of the firefighter job includes physical abilities, cognitive abilities and personal abilities.

In addition, the City has not established that the "weights" assigned to each of the nine cognitive abilities that Written Exams 7029 and 2043 were intended to measure reflect the importance of each ability to the job. As discussed throughout this report, the job analyses procedures used by the City were flawed, and the linking panel ratings are unreliable. Thus, there is no professional or scientific basis for the number of questions assigned each ability in these examinations.

Similar questions arise concerning the pass-fail cutoff scores. For Written Exam 7029, the City has admitted that the cutoff score was not based on data or information gathered from the job analysis. (Patitucci Dep., p. 93). Rather, the City deferred to the Fire Department, which requested that the cutoff score be based on the number of candidates (15,000) they wished to have pass. (*Id.* at p. 91). Indeed, the City's designee to testify about these matters indicated that he preferred a more "traditional" pass score (the score of 70 specified as a default cutoff score in the City's civil service regulations) because it would be more justifiable and defensible. (*Id.* at pp. 96-97). In any case, there is no professional or scientific basis for the setting of the cutoff score. For Written Exam 2043, DCAS assumed responsibility for the cutoff score and set it at 70, the civil service regulations' default cutoff score. (*Id.* at pp. 100-103). Again, there is no reference to professional or scientific standards in setting the cutoff score according to the civil service default score.

SECTION V.

CONCLUSION

In summary, it is my opinion that the City has not identified evidence that supports a claim of content validity for the City's use of Written Exams 7029 and 2043 in the selection and appointment of entry-level firefighters.